PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 9/00 | A2

(11) International Publication Number: WO 00/26349

(43) International Publication Date: 11 May 2000 (11.05.00)

(21) International Application Number: PCT/US99/24478

(22) International Filing Date: 22 October 1999 (22.10.99)

(30) Priority Data:

60/106,100	29 October 1998 (29.10.98)	US
60/120,254	16 February 1999 (16.02.99)	US

(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

(71) Applicant: KOSAN BIOSCIENCES, INC. [US/US]; 3832 Bay Center Drive, Hayward, CA 94545 (US).

(72) Inventors: BETLACH, Mary, C.; 2530 Diamond Street, San Francisco, CA 94131 (US). SHAH, Sanjay, Krishnakant; 4499 Sweetshrub Court, Concord, CA 94521 (US). MCDANIEL, Robert; 698 Matadero Avenue, Palo Alto, CA 94306 (US). TANG, Li; 1167-3 Foster City Boulevard, Foster City, CA 94404 (US).

(74) Agents: FAVORITO, Carolyn et al.; Morrison & Foerster LLP, 2000 Pennsylvania Avenue, N.W., Washington, DC 20006-1888 (US).

(54) Title: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE

(57) Abstract

Recombinant DNA compounds that encode all or a portion of the oleandolide polyketide synthase are used to express recombinant polyketide synthase genes in host cells for the production of oleandolide, oleandolide derivatives, and polyketides that are useful as antibiotics and motilides.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE

Field of the Invention

The present invention provides recombinant methods and materials for producing polyketides by recombinant DNA technology. The invention relates to the fields of agriculture, animal husbandry, chemistry, medicinal chemistry, medicine, molecular biology, pharmacology, and veterinary technology.

Background of the Invention

Polyketides represent a large family of diverse compounds synthesized from 2-carbon units through a series of condensations and subsequent modifications. Polyketides occur in many types of organisms, including fungi and mycelial bacteria, in particular, the actinomycetes. There are a wide variety of polyketide structures, and the class of polyketides encompasses numerous compounds with diverse activities. Erythromycin, FK-506, FK-520, narbomycin, oleandomycin, picromycin, rapamycin, spinocyn, and tylosin are examples of such compounds. Given the difficulty in producing polyketide compounds by traditional chemical methodology, and the typically low production of polyketides in wild-type cells, there has been considerable interest in finding improved or alternate means to produce polyketide compounds. See PCT publication Nos. WO 93/13663; WO 95/08548; WO 96/40968; 97/02358; and 98/27203; United States Patent Nos. 4,874,748; 5,063,155; 5,098,837; 5,149,639; 5,672,491; and 5,712,146; Fu *et al.*, 1994, *Biochemistry 33*: 9321-9326; McDaniel *et al.*, 1993, *Science 262*: 1546-1550; and Rohr, 1995, *Angew. Chem. Int. Ed. Engl. 34*(8): 881-888, each of which is incorporated herein by reference.

Polyketides are synthesized in nature by polyketide synthase (PKS) enzymes. These enzymes, which are complexes of multiple large proteins, are similar to the synthases that catalyze condensation of 2-carbon units in the biosynthesis of fatty acids. Two major types of PKS enzymes are known; these differ in their composition and mode of synthesis. These two major types of PKS enzymes are commonly referred to as Type I or "modular" and Type II "iterative" PKS enzymes.

Modular PKSs are responsible for producing a large number of 12-, 14-, and 16-membered macrolide antibiotics including erythromycin, methymycin, narbomycin, oleandomycin, picromycin, and tylosin. Modular PKS enzymes for 14-

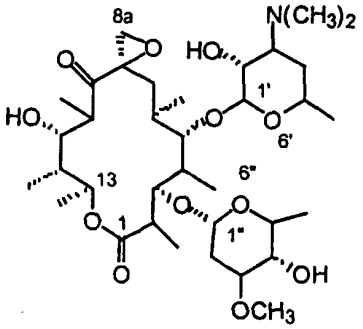
membered polyketides are encoded by PKS genes that often consist of three or more open reading frames (ORFs). Each ORF of a modular PKS can comprise one, two, or more "modules" of ketosynthase activity, each module of which consists of at least two (if a loading module) and more typically three (for the simplest extender module) or more enzymatic activities or "domains." These large multifunctional enzymes (>300,000 kDa) catalyze the biosynthesis of polyketide macrolactones through multistep pathways involving decarboxylative condensations between acyl thioesters followed by cycles of varying ß-carbon processing activities (see O'Hagan, D. *The polyketide metabolites*; E. Horwood: New York, 1991, incorporated herein by reference).

During the past half decade, the study of modular PKS function and specificity has been greatly facilitated by the plasmid-based *Streptomyces coelicolor* expression system developed with the 6-deoxyerythronolide B (6-dEB) synthase (DEBS) genes (see Kao *et al.*, 1994, *Science*, *265*: 509-512, McDaniel *et al.*, 1993, *Science 262*: 1546-1557, and U.S. Patent Nos. 5,672,491 and 5,712,146, each of which is incorporated herein by reference). The advantages to this plasmid-based genetic system for DEBS are that it overcomes the tedious and limited techniques for manipulating the natural DEBS host organism, *Saccharopolyspora erythraea*, allows more facile construction of recombinant PKSs, and reduces the complexity of PKS analysis by providing a "clean" host background. This system also expedited construction of the first combinatorial modular polyketide library in *Streptomyces* (see PCT publication No. WO 98/49315, incorporated herein by reference).

The ability to control aspects of polyketide biosynthesis, such as monomer selection and degree of ß-carbon processing, by genetic manipulation of PKSs has stimulated great interest in the combinatorial engineering of novel antibiotics (see Hutchinson, 1998, *Curr. Opin. Microbiol. 1*: 319-329; Carreras and Santi, 1998, *Curr. Opin. Biotech. 9*: 403-411; and U.S. Patent Nos. 5,712,146 and 5,672,491, each of which is incorporated herein by reference). This interest has resulted in the cloning, analysis, and manipulation by recombinant DNA technology of genes that encode PKS enzymes. The resulting technology allows one to manipulate a known PKS gene cluster either to produce the polyketide synthesized by that PKS at higher levels than occur in nature or in hosts that otherwise do not produce the polyketide. The

technology also allows one to produce molecules that are structurally related to, but distinct from, the polyketides produced from known PKS gene clusters.

Oleandomycin is an antibacterial polyketide (described in U.S. Patent No. 2,757,123, incorporated herein by reference) produced by a modular PKS in *Streptomyces antibioticus*. Oleandomycin has the structure shown below, with the conventional numbering scheme and stereochemical representation.



As is the case for certain other macrolide antibiotics, the macrolide product of the PKS , 8,8a-deoxyoleandolide, also referred to herein simply as oleandolide (although oleandolide in other contexts refers to the epoxidated aglycone), is further modified by epoxidation (at C-8 and C-8a) and glycosylation (an oleandrose at C-3 and a desosamine at C-5) to yield oleandomycin.

The reference Swan *et al.*, 1994, entitled "Characterisation of a *Streptomyces antibioticus* gene encoding a type I polyketide synthase which has an unusual coding sequence," *Mol. Gen. Genet. 242*: 358-362, incorporated herein by reference, describes the DNA sequence of the coding region of a gene designated *ORFB* hypothesized to encode modules 5 and 6 and a fragment of a gene designated *ORFA* hypothesized to contain the ACP domain of module 4 of the oleandolide PKS. The reference Quiros *et al.*, 1998, entitled "Two glycosyltransferases and a glycosidase are involved in oleandomycin modification during its biosynthesis by *Streptomyces antibioticus*," *Mol. Microbiol. 28*(6): 1177-1185, incorporated herein by reference, describes genes and gene products involved in oleandomycin modification during its biosynthesis. In particular, the reference describes a glycosyltransferase involved in rendering oleandomycin non-toxic to the producer cell and a glycosidase that reactivates oleandomycin after the glycosylated form is excreted from the cell. See also Olano *et al.*, Aug. 1998, "Analysis of a *Streptomyces antibioticus* chromosomal region involved in oleandomycin biosynthesis, which encodes two

glycosyltransferases responsible for glycosylation of the macrolactone ring, *Mol. Gen. Genet. 259*(3): 299-308, and PCT patent publication No. 99/05283, incorporated herein by reference. While a number of semi-synthetic oleandomycin derivatives have been described, see U.S. Patent Nos. 4,085,119; 4,090,017; 4,125,705; 4,133,950; 4,140,848; 4,166,901; 4,336,368; and 5,268,462, incorporated herein by reference, the number and diversity of such derivatives have been limited due to the inability to manipulate the PKS genes.

Genetic systems that allow rapid engineering of the oleandolide PKS would be valuable for creating novel compounds for pharmaceutical, agricultural, and veterinary applications. The production of such compounds could be accomplished if the heterologous expression of the oleandolide PKS in *Streptomyces coelicolor* and *S. lividans* and other host cells were possible. The present invention meets these and other needs.

Summary of the Invention

The present invention provides recombinant methods and materials for expressing PKS enzymes derived in whole and in part from the oleandolide PKS in recombinant host cells. The invention also provides the polyketides produced by such PKS enzymes. The invention provides in recombinant form all of the genes for the proteins that constitute the complete PKS that ultimately results, in *Streptomyces antibioticus*, in the production of oleandolide, which is further glycosylated and epoxidated to form oleandomycin. Thus, in one embodiment, the invention is directed to recombinant materials comprising nucleic acids with nucleotide sequences encoding at least one domain, module, or protein encoded by an oleandolide PKS gene. In one preferred embodiment of the invention, the DNA compounds of the invention comprise a coding sequence for at least one and preferably two or more of the domains of the loading module and extender modules 1 through 4, inclusive, of 8,8a-deoxyoleandolide synthase.

In one embodiment, the invention provides a recombinant expression vector that comprises a heterologous promoter positioned to drive expression of one or more of the oleandolide PKS genes. In a preferred embodiment, the promoter is derived from another PKS gene. In a related embodiment, the invention provides recombinant

host cells comprising the vector that produces oleandolide. In a preferred embodiment, the host cell is *Streptomyces lividans* or *S. coelicolor*.

In another embodiment, the invention provides a recombinant expression vector that comprises a promoter positioned to drive expression of a hybrid PKS comprising all or part of the oleandolide PKS and at least a part of a second PKS. In a related embodiment, the invention provides recombinant host cells comprising the vector that produces the hybrid PKS and its corresponding polyketide. In a preferred embodiment, the host cell is *Streptomyces lividans* or *S. coelicolor*.

In a related embodiment, the invention provides recombinant materials for the production of libraries of polyketides wherein the polyketide members of the library are synthesized by hybrid PKS enzymes of the invention. The resulting polyketides can be further modified to convert them to other useful compounds, such as antibiotics, typically through hydroxylation and/or glycosylation. Modified macrolides provided by the invention that are useful intermediates in the preparation of antibiotics are of particular benefit.

In another related embodiment, the invention provides a method to prepare a nucleic acid that encodes a modified PKS, which method comprises using the oleandolide PKS encoding sequence as a scaffold and modifying the portions of the nucleotide sequence that encode enzymatic activities, either by mutagenesis, inactivation, deletion, insertion, or replacement. The thus modified oleandolide PKS encoding nucleotide sequence can then be expressed in a suitable host cell and the cell employed to produce a polyketide different from that produced by the oleandolide PKS. In addition, portions of the oleandolide PKS coding sequence can be inserted into other PKS coding sequences to modify the products thereof.

In another related embodiment, the invention is directed to a multiplicity of cell colonies, constituting a library of colonies, wherein each colony of the library contains an expression vector for the production of a modular PKS derived in whole or in part from the oleandolide PKS. Thus, at least a portion of the modular PKS is identical to that found in the PKS that produces oleandolide and is identifiable as such. The derived portion can be prepared synthetically or directly from DNA derived from organisms that produce oleandolide. In addition, the invention provides methods to screen the resulting polyketide and antibiotic libraries.

The invention also provides novel polyketides, motilides, antibiotics, and other useful compounds derived therefrom. The compounds of the invention can also be used in the manufacture of another compound. In a preferred embodiment, the compounds of the invention are formulated as antibiotics in a mixture or solution for administration to an animal or human.

These and other embodiments of the invention are described in more detail in the following description, the examples, and claims set forth below.

Brief Description of the Figures

Figure 1 shows restriction site and function maps of the insert DNA in cosmids pKOS055-1 and pKOS055-5 of the invention. Various restriction sites (*Xho*I, *Cla*I, *Eco*RI) are also shown. Italicized restriction sites in the Figure indicate that not all of such sites are shown; the *Eco*RI sites shown are derived from the cosmid DNA into which the PKS gene segments were inserted. The location of the coding sequences for modules 1 – 6 of oleandolide PKS is indicated by brackets with labels underneath the brackets (i.e., mod. 2 is module 2). The sizes (in kilobase (kb) pairs) of various portions of the inserts are also shown. The open reading frames for the *oleAI* (*oleA1*), *oleAII* (*oleA2*), and *oleAIII* (*oleA3*) genes are shown as arrows pointing in the direction of transcription.

Figure 2 shows a function map of the oleandomycin gene cluster. In the top half of the Figure, the various open reading frames of the genes (*oleI, oleN2, oleR, oleAI, etc.*) are shown as arrows pointing in the direction of transcription. Directly beneath, a line indicates the size in base pairs (bp) of the gene cluster. The bar with alphanumeric identifiers under the size indicator line references Genbank accession numbers providing the nucleotide sequence of the indicated region, which sequence information is incorporated herein by reference. The cross-hatched portion of this bar indicates the region of the gene cluster for which sequence information is provided herein. In the bottom half of the Figure, the oleandolide PKS proteins are shown as arrow bars, with the location of the modules of the PKS shown below, and with the various domains of the modules shown below the modules.

Figure 3 shows a restriction site and function map of plasmid pKOS039-110, described in Example 3, below, which is an expression vector that can integrate (phiC31 based attachment and integration functions) into the chromosome of

Streptomyces and other host cells and contains the *ermE** promoter positioned to drive expression of the *oleAI* gene.

Figure 4 shows a restriction site and function map of plasmid pKOS039-130, described in Example 4, below, which is an expression vector that replicates (SCP2* origin of replication) in *Streptomyces* host cells and contains the *actI* promoter and *actII-ORF4* activator positioned to drive expression of the *oleAI*, *oleAII*, and *oleAIII* genes.

Figure 5 shows a restriction site and function map of plasmid pKOS039-133, described in Example 5, below, which is an expression vector that can integrate (phiC31 based attachment and integration functions) into the chromosome of *Streptomyces* and other host cells and contains the *actI* promoter and *actII-ORF4* activator positioned to drive expression of the *oleAIII* gene.

Detailed Description of the Invention

The present invention provides useful compounds and methods for producing polyketides in recombinant host cells. As used herein, the term recombinant refers to a compound or composition produced by human intervention. The invention provides recombinant DNA compounds encoding all or a portion of the oleandolide PKS. The invention provides recombinant expression vectors useful in producing the oleandolide PKS and hybrid PKSs composed of a portion of the oleandolide PKS in recombinant host cells. The invention provides the polyketides produced by the recombinant PKS as well as those derived therefrom by chemical processes and/or by treatment with polyketide modification enzymes.

To appreciate the many and diverse benefits and applications of the invention, the description of the invention below is organized as follows. In Section I, the recombinant oleandolide PKS provided by the invention is described. In Section II, methods for heterologous expression of the oleandolide PKS and oleandolide modification enzymes provided by the invention are described. In Section III, the hybrid PKS genes provided by the invention and the polyketides produced thereby are described. In Section IV, the polyketide compounds provided by the invention and pharmaceutical compositions of those compounds are described. The detailed description is followed by a variety of working examples illustrating the invention.

The oleandolide synthase gene, like other PKS genes, is composed of coding sequences organized in a loading module, a number of extender modules, and a thioesterase domain. As described more fully below, each of these domains and modules is a polypeptide with one or more specific functions. Generally, the loading module is responsible for binding the first building block used to synthesize the polyketide and transferring it to the first extender module. The building blocks used to form complex polyketides are typically acylthioesters, most commonly acetyl, propionyl, malonyl, 2-hydroxymalonyl, 2-methylmalonyl, and 2-ethylmalonyl CoA. Other building blocks include amino acid like acylthioesters. PKSs catalyze the biosynthesis of polyketides through repeated, decarboxylative Claisen condensations between the acylthioester building blocks. Each module is responsible for binding a building block, performing one or more functions, and transferring the resulting compound to the next module. The next module, in turn, is responsible for attaching the next building block and transferring the growing compound to the next module until synthesis is complete. At that point, an enzymatic thioesterase activity cleaves the polyketide from the PKS.

Such modular organization is characteristic of the class of PKS enzymes that synthesize complex polyketides and is well known in the art. The polyketide known as 6-deoxyerythronolide B (6-dEB) is a classic example of this type of complex polyketide. The genes, known as *eryAI*, *eryAII*, and *eryAIII* (also referred to herein as the DEBS genes, for the proteins, known as DEBS1, DEBS2, and DEBS3, that comprise the 6-dEB synthase), that code for the multi-subunit protein known as DEBS that synthesizes 6-dEB are described in U.S. Patent No. 5,824,513, incorporated herein by reference. Recombinant methods for manipulating modular PKS genes are described in U.S. Patent Nos. 5,672,491; 5,843,718; 5,830,750; and 5,712,146; and in PCT publication Nos. 98/49315 and 97/02358, each of which is incorporated herein by reference.

The loading module of DEBS consists of two domains, an acyl-transferase (AT) domain and an acyl carrier protein (ACP) domain. Each extender module of DEBS, like those of other modular PKS enzymes, contains a ketosynthase (KS), AT, and ACP domains, and zero, one, two, or three domains for enzymatic activities that modify the beta-carbon of the growing polyketide chain. A module can also contain domains for other enzymatic activities, such as, for example, a methyltransferase

activity. Finally, the releasing domain contains a thioesterase and, often, a cyclase activity.

The AT domain of the loading module recognizes a particular acyl-CoA (for DEBS this is usually propionyl but sometimes butyryl or acetyl) and transfers it as a thiol ester to the ACP of the loading module. Concurrently, the AT on each of the extender modules recognizes a particular extender-CoA (malonyl or alpha-substituted malonyl, i.e., methylmalonyl, ethylmalonyl, and 2-hydroxymalonyl) and transfers it to the ACP of that module to form a thioester. Once the PKS is primed with acyl- and malonyl-ACPs, the acyl group of the loading module migrates to form a thiol ester (trans-esterification) at the KS of the first extender module; at this stage, extender module 1 possesses an acyl-KS and a malonyl (or substituted malonyl) ACP. The acyl group derived from the loading module is then covalently attached to the alpha-carbon of the malonyl group to form a carbon-carbon bond, driven by concomitant decarboxylation, and generating a new acyl-ACP that has a backbone two carbons longer than the loading unit (elongation or extension). The growing polyketide chain is transferred from the ACP to the KS of the next module, and the process continues.

The polyketide chain, growing by two carbons each module, is sequentially passed as a covalently bound thiol ester from module to module, in an assembly line-like process. The carbon chain produced by this process alone would possess a ketone at every other carbon atom, producing a polyketone, from which the name polyketide arises. Most commonly, however, additional enzymatic activities modify the beta keto group of each two-carbon unit just after it has been added to the growing polyketide chain but before it is transferred to the next module. Thus, in addition to the minimal module containing KS, AT, and ACP domains necessary to form the carbon-carbon bond, modules may contain a ketodreductase (KR) that reduces the keto group to an alcohol. Modules may also contain a KR plus a dehydratase (DH) that dehydrates the alcohol to a double bond. Modules may also contain a KR, a DH, and an enoylreductase (ER) that converts the double bond to a saturated single bond using the beta carbon as a methylene function. As noted above, modules may contain additional enzymatic activities as well.

Once a polyketide chain traverses the final extender module of a PKS, it encounters the releasing domain or thioesterase found at the carboxyl end of most PKSs. Here, the polyketide is cleaved from the enzyme and cyclyzed. The resulting

polyketide can be modified further by tailoring or polyketide modification enzymes; these enzymes add carbohydrate groups or methyl groups, or make other modifications, i.e., oxidation or reduction, on the polyketide core molecule.

While the above description applies generally to modular PKS enzymes, there are a number of variations that exist in nature. For example, some polyketides, such as epothilone, incorporate a building block that is derived from an amino acid. PKS enzymes for such polyketides include an activity that functions as an amino acid ligase or as a non-ribosomal peptide synthetase (NRPS). Another example of a variation, which is actually found more often than the two domain loading module construct found in DEBS, occurs when the loading module of the PKS is not composed of an AT and an ACP but instead utilizes an inactivated KS, an AT, and an ACP. This inactivated KS is in most instances called KS^Q, where the superscript letter is the abbreviation for the amino acid, glutamine, that is present instead of the active site cysteine required for activity. For example, the oleandolide PKS loading module contains a KS^Q. Yet another example of a variation has been mentioned above in the context of modules that include a methyltransferase activity; modules can also include an epimerase activity. The components of a PKS are described further below in specific reference to the oleandolide PKS and the various recombinant and hybrid PKSs provided by the invention.

Section I: The Oleandolide PKS

The oleandolide PKS was isolated and cloned by the following procedure. Genomic DNA was isolated from an oleandomycin producing strain of *Streptomyces antibioticus* (ATCC 11891), partially digested with a restriction enzyme, and cloned into a commercially available cosmid vector to produce a genomic library. This library was then introduced into *E. coli* and probed with a DNA fragment generated from *S. antibioticus* DNA using primers complementary to sequences of KS domains encoding extender modules 5 and 6 of the oleandolide PKS. Several colonies that hybridized to the probe were pooled, replated, and probed again, resulting in the identification of a set of cosmids. These latter cosmids were isolated and transformed into a commercially available *E. coli* strain. Plasmid DNA was isolated and analyzed by DNA sequence analysis and restriction enzyme digestion, which revealed that the

desired DNA had been isolated and that the entire PKS gene cluster was contained in overlapping segments on two of the cosmids identified.

Further analysis of these cosmids and subclones prepared from the cosmids facilitated the identification of the location of various oleandolide PKS genes and ORFs, as well as the modules and domains in the PKS proteins encoded by those ORFs. The location of these genes and modules is shown on Figures 1 and 2. Figure 1 shows that the complete oleandolide PKS gene cluster is contained within the insert DNA of cosmids pKOS055-1 (insert size of ~43 kb) and pKOS055-5 (insert size of ~47 kb). Each of these cosmids has been deposited with the American Type Culture Collection in accordance with the terms of the Budapest Treaty (cosmid pKOS055-1 is available under accession no. ATCC 203798; cosmid pKOS055-5 is available under accession no. ATCC 203799). Various additional reagents of the invention can be isolated from these cosmids. DNA sequence analysis was also performed on the various subclones of the invention, as described herein.

Those of skill in the art will recognize that, due to the degenerate nature of the genetic code, a variety of DNA compounds differing in their nucleotide sequences can be used to encode a given amino acid sequence of the invention. The native DNA sequence encoding the oleandolide PKS of *Streptomyces antibioticus* is shown herein merely to illustrate a preferred embodiment of the invention, and the invention includes DNA compounds of any sequence that encode the amino acid sequences of the polypeptides and proteins of the invention. In similar fashion, a polypeptide can typically tolerate one or more amino acid substitutions, deletions, and insertions in its amino acid sequence without loss or significant loss of a desired activity. The present invention includes such polypeptides with alternate amino acid sequences, and the amino acid sequences encoded by the DNA sequences shown herein merely illustrate preferred embodiments of the invention.

The recombinant nucleic acids, proteins, and peptides of the invention are many and diverse. To facilitate an understanding of the invention and the diverse compounds and methods provided thereby, the following description of the various regions of the oleandolide PKS and corresponding coding sequences is provided. To facilitate description of the invention, reference to a PKS, protein, module, or domain herein can also refer to DNA compounds comprising coding sequences therefor and *vice versa*. Also, unless otherwise indicated, reference to a heterologous PKS refers to

a PKS or DNA compounds comprising coding sequences therefor from an organism other than *Streptomyces antibioticus*. In addition, reference to a PKS or its coding sequence includes reference to any portion thereof.

Thus, the invention provides DNA molecules in isolated (i.e., not pure, but existing in a preparation in an abundance and/or concentration not found in nature) and purified (i.e., substantially free of contaminating materials or substantially free of materials with which the corresponding DNA would be found in nature) form. These DNA molecules comprise one or more sequences that encode one or more domains (or fragments of such domains) of one or more modules in one or more of the ORFs of the oleandolide PKS gene cluster. Examples of such domains include the KS, AT, DH, KR, ER, ACP, and TE domains of at least one of the 6 extender modules and loading module encoded by the 3 ORFs of the oleandomycin PKS genes.

In one embodiment, the DNA molecule comprises an ORF other than or in addition to the ORFB described in Swan *et al.*, *supra*; which corresponds to the *oleAIII* gene ORF herein, the module is a module other than or in addition to extender module 5 and/or module 6 of *ORFB*; and the domain is a domain other than or in addition to a domain of module 5 and/or module 6 of *ORFB* or the ACP domain of module 4 of *ORFA*. In an especially preferred embodiment, the DNA molecule is a recombinant DNA expression vector or plasmid. Such vectors can either replicate in the cytoplasm of the host cell or integrate into the chromosomal DNA of the host cell. In either case, the vector can be a stable vector (i.e., the vector remains present over many cell divisions, even if only with selective pressure) or a transient vector (i.e., the vector is gradually lost by host cells with increasing numbers of cell divisions).

The oleandolide PKS, also known as 8, 8a-deoxyoleandolide synthase, is encoded by three ORFs (*oleAI*, *oleAII*, and *oleAIII*). Each ORF encodes 2 extender modules of the PKS; the first ORF also encodes the loading module. Each module is composed of at least a KS, an AT, and an ACP domain. The locations of the various encoding regions of these ORFs are shown in Figure 2 and described with reference to the sequence information below.

ORF1 encodes 8, 8a-deoxyoleandolide synthase I and begins at nucleotide 5772 and ends at nucleotide 18224 in the sequence below. ORF1 encodes a loading module (encoded by nucleotides 5799-8873), composed of a KS^Q domain (encoded by nucleotides 5799-7055), a malonyl-specific AT domain (encoded by nucleotides

7458-8563), and an ACP domain (encoded by nucleotides 8634-8873). ORF1 also encodes extender module 1 (encoded by nucleotides 8955-13349), composed of a KS domain (KS1, encoded by nucleotides 8955-10205), an AT domain (AT1, encoded by nucleotides 10512-11549), a KR domain (KR1, encoded by nucleotides 12258-12818), and an ACP domain (ACP1, encoded by nucleotides 13092-13349), and extender module 2 (encoded by nucleotides 13407-17966), composed of a KS domain (KS2, encoded by nucleotides 13407-14690), an AT domain (AT2, encoded by nucleotides 14997-16031), a KR domain (KR2, encoded by nucleotides 16872-17423), and an ACP domain (ACP2, encoded by nucleotides 17709-17996).

ORF2 encodes 8, 8a-deoxyoleandolide synthase 2 and begins at nucleotide 18267 and ends at nucleotide 29717 in the sequence below. ORF2 encodes extender module 3 (encoded by nucleotides 18357-22985), composed of a KS domain (KS3, encoded by nucleotides 18357-19643), an AT domain (AT3, encoded by nucleotides 19965-20999), an inactive KR domain (KR3, encoded by nucleotides 21897-22449), and an ACP domain (ACP3, encoded by nucleotides 22728-22985), and extender module 4 (encoded by nucleotides 23046-29396), composed of a KS domain (KS4, encoded by nucleotides 23046-24329), an AT domain (AT4, encoded by nucleotides 24645-25682), a DH domain (DH4, encoded by nucleotides 25719-26256), an ER domain (ER4, encoded by nucleotides 27429-28301), a KR domain (KR4, encoded by nucleotides 28314-28862), and an ACP domain (ACP4, encoded by nucleotides 29147-29396).

ORF3 encodes 8, 8a-deoxyoleandolide synthase 3 and begins at nucleotide 29787 and ends at nucleotide 40346 in the sequence below. This sequence has been previously reported by Swan *et al.*, *supra*. ORF3 encodes extender module 5 (encoded by nucleotides 29886-34478), composed of a KS domain (KS5, encoded by nucleotides 29886-31184), an AT domain (AT5, encoded by nucleotides 31494-32531), a KR domain (KR5, encoded by nucleotides 33384-33935), and an ACP domain (ACP5, encoded by nucleotides 34221-34478), and extender module 6 (encoded by nucleotides 34845-39440), composed of a KS domain (KS6, encoded by nucleotides 34845-36131), an AT domain (AT6, encoded by nucleotides 36447-37484), a KR domain (KR6, encoded by nucleotides 38352-38903), and an ACP domain (ACP6, encoded by nucleotides 39183-39440). ORF3 also encodes a TE domain at nucleotides 39657-40343.

The DNA sequence below also includes the sequences of a number of the tailoring enzyme genes in the oleandomycin gene cluster, including *oleI* (nucleotides 152-1426), *oleN2* (nucleotides 1528-2637), *oleR* (nucleotides 2658-4967), *olePI* (nucleotides 40625-41830), *oleG1* (nucleotides 41878-43158), *oleG2* (nucleotides 43163-44443), *oleM1* (nucleotides 44433-45173), *oleY* (nucleotides 45251-46411), *oleP* (nucleotides 46491-47714), and *oleB* (nucleotides 47808-49517).

The sequence of the portion of the oleandomycin gene cluster described above follows:

```
   1 GCATGCCCGC CCGCAACACC GGCTCCCGTA ACGGGGCGAG CCGGTGGTCA TCCATCAGTT
  61 TCCTTCCGCC CGGCCCGTGT CAGGCCCGTG TGCGCATACC GCCGTACGGC TGCGCCGGTC
 121 CCCCGCGGAA CACCTCACCG GAGTGAGATC CATGACGAGC GAGCACCGCT CTGCCTCCGT
 181 GACACCCCGT CACATCTCCT TCTTCAACAT CCCCGGCCAC GGCCACGTGA ACCCGTCACT
 241 CGGCATTGTC CAGGGACTTG TCGCGCGCGG CCAACGGGTC AGCTACGGCA TTACCGACGA
 301 GTTCGGCGCA CAGGTCAAGG CGGGCCGCGC GACGGCCGTT GTGTACGGCT TCATTCTGCC
 361 GGAGGAGTTC AACCCCGAGG AGTTGTTGGC CGAGGACCAG GGTTCCCGAT GGGCCTGTTC
 421 CTTGGCGGAG GCGTTCCGGG TCTTGCCGCA GCTGAGGACG GCTACGCCGA CGACCGGCCG
 481 GGACCTGATC GTCTACGACA TCGCCTCCTG GCCCGCCCCG GTGCTCGGCC GGAAGTGGGA
 541 CATCCCCTTC GTCCAGCTCT CCCCGACCTC CGTCGCCTAC GAGGGCTTCG AGGAGGACGT
 601 ACCCGCGGTG CAGGACCCCA CGGCCGACCG CGGCGAGGAG GCCGCCGCCC CCGCGGGGAC
 661 CGGGGACGCC GAGGAGGGTG CCGAGGCCGA GGACGGCCTG GTGCGCTTCT TCACCCGGCT
 721 CTCGGCCTTC CTGGAGGAGC ACGGGGTGGA CACCCCGGCC ACCGAGTTCC TCATCGCGCC
 781 CAACCGCTGC ATCGTCGGCT GCCGCGCACC TTCCCAGATC AAGGGCGACA CGGTCGGCGA
 841 CAACTACACC TTCGTCGGTC CCACCTACGG CGACCGGTCC CACCAGGGCA CCTGGGAAGG
 901 CCCCGGGCAC GGGCGTCCGG TGCTGCTGAT CGCCCTGGGC TCGGCGTTCA CCGACCACCT
 961 CGACTTCTAC CGCACCTGCC TGTCCGCCGT CGACGGCCTG GACTGGCACG TGGTGCTCTC
1021 CGTGGGCCGC TTCGTCGACC CCGCGGACCT CGGCGAGGTC CCGCCGAACG TCGAGGTGCA
1081 CCAGTGGGTG CCGCAGCTCG ACATCCTGAC CAAAGCCTCC GCGTTCATCA CGCACGCGGG
1141 CATGGGCAGC ACCATGGAGG CCCTGTCGAA CGCGGTGCCC ATGGTCGCGG TGCCGCAGAT
1201 CGCGGAGCAG ACGATGAACG CCGAGCGGAT CGTCGAGCTG GGCCTCGGCC GGCACATCCC
1261 GCGGGACCAG GTCACGGCCG AGAAGCTGCG CGAGGCCGTG CTCGCCGTCG CCTCCGACCC
1321 CGGTGTCGCC GAACGGCTCG CGGCCGTCCG GCAGGAGATC CGTGAGGCGG GCGGCGCCCG
1381 GGCGGCCGCC GACATCCTGG AGGGCATCCT CGCCGAAGCA GGCTGACCGC CCCTGCCTGA
1441 CGGTGCGCGG GCCGCCCGGC CCGCCGCGTG AGAGTCGGCC CCCGTACCCG ACGACGGGTA
1501 CGGGGGCCGA CGCGCGCGGG CCCGGACTCA GCAGGCGGCC ACCGCGCCCC GTACCGCCTC
1561 GATCACCGCC TTGACGGCGT CGTCGGACAG GTGCGGGCCT ATGGGCAGGC TCAGCACCTC
1621 CCGGGCGAGC CGCTCCGCCA CGGGCTGTGC GCGGGCGGCC TGCCGGCTGC CGGCGTACGC
1681 CTCCGACCGG TGCACCGGCA CCGGGTAGTG GATCAGCGTC TCGACGCCGG CTGCCGCCAG
1741 CCGCTCCCGC AGCGCGGACC GGTCCGCGGA ACGAATCACG AACAGGTGCC ACACGGGGTC
1801 CGCCCACGGC GCCGGCCTCG GCAGCACGAT CCCGTCCAGG CCGGCGAGCC CGTCGAGATA
1861 GCGCGCCGCC ACCGCGGCCC GGCGCTCGGG TCCCAGCCGT CCCAGGTGGG CGAGCTTGAC
1921 CCGCAGAACG GCCGCTTGCA GCTCGTCCAG CCGGAAGTTG GTGGCCCGGA CCTCGTGCCG
1981 GTACTTCTCC CGCGACCCGT AGTTGCGCAG CAGCCGCACC CGCTCCGCCA GCTCCGCGTC
2041 GTCCGTCACC ACGGCGCCGC CGTCACCGAA GCCGCCCAGG TTCTTGCCCG GGTAGAAGCT
2101 GAAGGCGGTG GTGGACCACG CGCCCACCCG CCGGCCGTAC GCCTGCGCAC CGTGCGCCTG
2161 GGCGGCGTCC TCCAGGATCC GCACGCCGTG CCGCTCGGCG ACCTCGGACA ACGCCGCCAG
2221 GTCCGCCGGA TGCCCGTACA GGTGCACCGG GAGGATCACC CGGGTGCGGG AGGTGATCGC
2281 AGCCTCGACG CGCTCCGGGT CCAGGGTGAA CGTCGCAGGC TCCGGTTCCA CCGCGACGGG
2341 CTCCGCACCC GTCGCCGAGA CGGCGAGCCA GGTCGCGGCG AAGGTGTGCG CCGGGACGAT
2401 CACCTCGTCA CCCGGCCCGA TGTCCATGGC GCGCAGCGCC AGTTCCAGGG CGTCGCACCC
2461 GCTGCCCACC GCCACGCAGT GCCGGGCCCC GCAGTAGGCG GCCCACTCCG TCTCGAACGC
2521 GGCGAGTTCG GGGCCCAGGA GGTAGCGCCC GGAGTCCAGG ACGCGGCCGG TCGCGGCGTC
2581 GATGTCGTGC TTGAGCTCCA GGTAGGCGGC CCGGAGGTCC AGGAACGGAA CGTCCATGCG
2641 TCCTCCGTGG GAGCTGCTCA CGGCGCCGTG GCGCTGAGCG GGAGACGGCC GAGGGACGGG
2701 CCCACCATGA CCTGCCGTCC GGGTCCGGTC ACCCAGGTGT GGGCGCCGCT GTCCCAGTTC
2761 TGGAGGGCCC TGCGCTCGAC GTGCAGGGTC AGCCTCCTGC TCTCGCCCGG CCGCAGCTCG
2821 ACCTTCCCGT AGGCCGCCAG GGCACGCTTG GCCTGCGCCA CCCGCACGTG CGGGGACGGC
2881 CCCACGTAGA CCTGCGGGAC CTCCTTGCCG GTGCGCGTAC CGGTGTTGCG CAGCGTGAAG
2941 CAGACGTCGA GCCCGCCGTC CGCCGTCGCC GTCACCTTCA GGTCCCGGTA GTCGAAGGAG
3001 GTGTAGCACA ACCCGTGGCC GAAGGAGAAC AGCGGCTGGA CGCCCTGCTG TTCGTACCAG
```

```
3061 CGGTAGCCGG AGTAGATGCC CTCGGAGTAG TCCAGTTGGT CATCGACTCC CGGGTAGCGC
3121 CTGGCGTCCC CGGCGAACGG CGTCTGCCCC TCGTCGGCCG GGAAGGTCTG GGTCAGCCGG
3181 CCTCCTGGGT CGGCGTCGCC GAACAGCAGG GCGGTGGTCG CCTCGGCGCC GGCCTGGCCC
3241 GGGTACCACA TGGTGAGCAC CGCGGCGGTC TTCCTCAGCC AGGGCATGGT GAGGGAGGAG
3301 CCCGTGTTGA GCACCACCAC GGTCCGTGGG TTGACCGCGG CCACGGCGCT GATCAGGTCG
3361 TCCTGGCGGC CGGGCAGGGA CAGCGACGTG CGGTCCCCGT CCTCCGAGCC GTCGTCGTAC
3421 GCGAAGACGA CCGCGGTCCT CGCCGTCCGC GCGATCGACA CGGCCCGGTC GATCGCCTCC
3481 TGGGCGGCCT GCGGAGTGAC CCACGTCAGC TCGAAGGTCA TGGGCGACTT CGCCAGGGCC
3541 GCGCCGGTGA TGCGCAGCTT GTGCGTTCCG GCCGCCAGCC GCATGGGGCG GCTGCTGACG
3601 TCGCCGTAGA CCCAGGGCCG ACGGCCGAAC GGCTCCTGGC CGTCGAGTTC GACGTAGGCG
3661 TTGCCGCCCT GCGCGCGGGC CGCGATGCGG TAGCTGCCGG TGACCGGCAC GGTGATGGTG
3721 CCGTCGTAGA GGACACCGCC CCCACCGGCG GGGAACACCT CGCCCGAGGG GCGGGGCCGC
3781 GGAAGAGGGG CGGACTGCGG AACGGGAACC CCGACCGTCT CCTCACCGGT GCTGTAGCGC
3841 ACGGTGCTGC CGGCGCCGGC CCGTTCGCGG ATGGTGTCCA GAGGGGCGGA CGCGCCGTCC
3901 GGCACGATGT ACGAACTGCC CAGCCCGGTC ACCTTCGGGA CCTTGGCGGT GGGGCCGATC
3961 ACGGCGATGT CCGCCGCCGT CTCCGTGGTC AGGGGAAGGG TGGCGCCCTC GTTGCGCAGC
4021 AGGACCGCGC CGTCCTCGGC GACCTGGCGC GCGACCTTCA AGCCGCCCGC GAGGTCGCGC
4081 GCCGGGCGGG CGGGCGGATC CTCGTCCAGC AGCCGGAACC GGGCCATCTG CGACACGATG
4141 CGGGTGACGG CCTCGTCGAG GGCCGACTCG GGGATGCGTC CCTCCCGGAT CGCCGTCTTG
4201 AGCGGGTCGC CGAAGAACTT GCCGCCGGGT ATCGGCTCGC CCGGGGCGGG TTCGTGGTCC
4261 AGCTCGATGC CGAGTTCCTG GTCGAGCCCC TTGGTGAGGG CGTCCGTGCT CTGCGTCGCC
4321 AGCCAGTCCG AGGTCACCCA GCCACGGAAC TTCCACTGCT CCTTGAGGAC CTTGTTCAGC
4381 AGTTCGTCAC TGCCGCAGGC CGGCTGGCCG TTGACCTTGT TGTAGGCGCA CATCACCGAG
4441 CCGGTTCCGG CAGCCACGGC GCTCTCGAAA CCGGGCAGTT CCCGCTCGCG CAACGTCTGT
4501 TCGTCGACGT TCACGTTAAC GCTGAAACGA TTCTTCTCCT GGTTGTTCGC CGCGTAGTGC
4561 TTGGTGGCGG CGATCAGCCC CTGACTCTGG ATGCCCTTGA TCTCCGCGGC GGCCATCCGC
4621 GAGGTGACCA GGGGGTCCTC GCTGAACGTC TCGAAGTTCC GCCCGGCGTA CGGCACGCGT
4681 ATGGAGTTCA CCATCGGCGC GAACACCACG TCCTGCCCGA AGGCGCGCCC CTCCCGGCCG
4741 ATCACCGCCC CGTAGGACCG CGCCAGGCCG TCGTCGAAGG TGGAGGCCAG CGCCACGGGA
4801 GCGGGCAGCG CGAGGGACGG CCGGTGGATC GTGATTCCGG CGGGACCGTC GGTGGCCCGC
4861 ATCTCGGGTA TGCCGAGGCG GGGAACGCCC GGCAGGTACA CCTTTGCCGA CTCATCGCTC
4921 GTGTGATAGC TCCAGTGCAC GAACGACAGC TTTTCTTCCA GGGTCATCCG AGCCGTCAGA
4981 AGACGAGCCG TTTCCCACGG ATCGCCCGAT TCGGCGACGG ACGGAACAGA GGGGAGCAGG
5041 GCGAGACCGA GGGCCAGGCC GAGAGTACCC GCGGAGGTCC GTGGCGGGAC CGGACTCCTG
5101 CGCTGCGCAC GGCCGCCGAG ACGTAACCGA AGTGATCTCA AAAGGCTTCC AAATCCTCCG
5161 CGCCCTCGTG CTGCGAGGCG CATGAAATGG GCGGTTGTCG CGACCACAGT GCACCGTCAC
5221 CGAAGCCGGA GCAATGCCCG TGAATAAGGT CGCGCCCTTC CGTGGATGAT CTCCGCACGA
5281 GATCATGCCC AGCTCAAGTG ATGGTCATGC ACGTACCAAG AAGGGGCTTG CCTGGGGGGC
5341 GTGAGCTGAT CTAGCGTTGC CGCACGACGA CGAGTCGTGA GCGAGGCGAA CGCTCTGCCG
5401 CTCAGGGGGT GAACAGACGG CAGCCCGGAC GTTCGACGAG GGTCAAGCGG AACGCAGGCG
5461 ACAGGACGCG GCCACCCTCC GAGGCACCCG TGCCGACCAT CCTCGCAGGT CCTTCGCCAT
5521 GCCCGTCGCA ACTCTCCGAT CGCTGCCGCC GATGGCGACA GCCCGGCACC GAGGCCCCTG
5581 GACCAGGAGG CGAAGCGAGG GCCGGCCGCG ATGCACGAAT CGGACCCAGG CGAACACCGG
5641 CACATCCACC CCGGCGCGTG CGGTACGGGC CGCGCCCGAT GACGGGCGAA CGACGACCGA
5701 AAAGCAGACC CCTTGATTCG CTTCCATGGT TGTGGCAGCC GCGGGGAGCG TCGGCAGAGA
5761 GGTGGGAAAC CATGCATGTC CCCGGCGAGG AAAACGGGCA TTCCATTGCC ATTGTCGGAA
5821 TTGCGTGCCG ACTGCCGGGC TCTGCCACCC CCCAGGAGTT CTGGAGACTC CTGGCCGACT
5881 CCGCAGACGC ATTGGACGAG CCCCCCGCCG GCCGTTTCCC GACCGGCTCA TTATCCTCGC
5941 CCCCCGCTCC GCGCGGCGGA TTCCTCGACA GCATCGACAC TTTCGACGCG GATTTCTTCA
6001 ACATCTCGCC CAGAGAAGCC GGTGTCCTCG ACCCCCAGCA ACGCCTCGCG CTGGAACTCG
6061 GCTGGGAGGC GCTGGAAGAC GCCGGAATCG TCCCGCGACA CCTCAGGGGA ACCCGCACCT
6121 CGGTCTTCAT GGGCGCCATG TGGGACGACT ACGCGCACCT GGCGCACGCA CGGGGAGAAG
6181 CCGCCCTCAC CCGGCATTCC CTGACGGGAA CGCACCGCGG CATGATCGCC AACCGGCTCT
6241 CCTACGCCCT GGGCCTCCAA GGCCCCAGCC TCACCGTCGA CACCGGACAA TCCTCCTCCC
6301 TCGCCGCCGT GCACATGGCC TGCGAGAGCC TGGCCCGCGG CGAATCCGAC CTCGCCCTCG
6361 TCGGCGGCGT CAACCTCGTC CTCGATCCGG CCGGCACGAC CGGCGTCGAG AGGTTCGGAG
6421 CACTCTCACC GGACGGCAGG TGCTACACCT TCGACTCCCG GGCGAACGGC TACGCCCGAG
6481 GAGAGGGCGG CGTCGTAGTC GTCCTCAAGC CCACCCACCG CGCGCTCGCG GACGGTGACA
6541 CCGTCTACTG CGAGATCCTG GGCAGCGCCC TCAACAACGA CGGCGCCACG GAAGGCCTCA
6601 CCGTCCCCAG CGCCCGCGCC CAGGCGGACG TCCTGCGACA GGCATGGGAA CGGGCACGCG
6661 TGGCCCCGAC GGACGTCCAG TACGTGGAAC TGCACGGAAC CGGCACACCG GCCGGCGACC
6721 CCGTCGAGGC CGAGGGCCTC GGCACCGCGC TCGGCACCGC ACGCCCGGCC GAGGCGCCGC
6781 TCCTGGTCGG CTCGGTCAAG ACGAACATCG GTCACCTCGA AGGCGCGGCA GGCATCGCGG
6841 GCCTCCTGAA GACGGTCCTG AGCATCAAGA ACCGGCACCT CCCGGCAAGC CTGAACTTCA
6901 CCTCGCCCAA CCCCCGCATC GACCTCGACG CCCTGCGCCT GCGCGTCCAC ACCGCGTACG
6961 GCCCCTGGCC GAGCCCCGAC CGGCCGCTGG TGGCGGGCGT CTCCTCCTTC GGCATGGGCG
7021 GGACGAACTG CCACGTCGTC CTGTCCGAGT TACGGAACGC GGGAGGCGAC GGCGCCGGAA
```

```
 7081 AAGGGCCGTA CACCGGCACG GAAGACCGGC TCGGCGCCAC GGAGGCGGAG AAGAGGCCGG
 7141 ACCCGGCAAC CGGAAACGGT CCTGATCCCG CCCAGGACAC CCACCGCTAC CCGCCGCTGA
 7201 TCCTGTCCGC CCGCAGCGAC GCGGCCCTGC GCGCACAGGC GGAACGGCTC CGCCACCACC
 7261 TGGAACACAG CCCCGGACAG CGCCTGCGGG ACACCGCCTA CAGCCTGGCG ACCCGCCGCC
 7321 AGGTCTTCGA GCGGCACGCG GTGGTCACCG GACACGACCG CGAGGACCTG CTCAACGGCC
 7381 TGCGTGACCT GGAGAACGGC CTCCCGGCCC CCCAGGTCCT GCTCGGCCGC ACGCCCACCC
 7441 CCGAACCGGG CGGCCTCGCC TTCCTCTTCT CCGGGCAGGG CAGCCAGCAG CCCGGCATGG
 7501 GCAAGCGACT CCACCAGGTG TTCCCCGGCT TCCGGGACGC CCTGGACGAG GTCTGCGCCG
 7561 AACTCGACAC CCACCTCGGC CGACTCCTCG GCCCCGAGGC CGGCCCGCCC CTGCGCGACG
 7621 TGATGTTCGC CGAGCGGGGC ACGGCGCACA GCGCCCTGCT CTCCGAGACC CACTACACCC
 7681 AGGCCGCCCT CTTCGCCCTG GAAACCGCCC TCTTCCGCCT CCTGGTCCAG TGGGGCCTGA
 7741 AACCCGACCA CCTCGCAGGC CACTCCGTCG GCGAGATCGC GGCCGCCCAC GCAGCAGGCA
 7801 TCCTCGACCT GTCCGACGCG GCCGAACTCG TGGCCACCCG CGGCGCGTTG ATGCGTTCCC
 7861 TGCCCGGCGG CGGCGTCATG CTCTCGGTCC AGGCACCCGA GTCCGAGGTC GCACCCCTGC
 7921 TGCTCGGCCG TGAGGCCCAC GTCGGCCTGG CCGCCGTGAA CGGCCCCGAC GCGGTGGTCG
 7981 TGTCCGGCGA GCGCGGCCAC GTCGCCGCCA TCGAACAGAT CCTCCGGGAC AGGGGCCGCA
 8041 AAAGCCGGTA CCTGCGCGTC AGCCACGCCT TCCACTCCCC GCTCATGGAA CCGGTGCTGG
 8101 AGGAGTTCGC CGAAGCCGTC GCCGGCCTGA CCTTCCGGGC ACCGACCACA CCCCTCGTCT
 8161 CCAACCTCAC CGGCGCACCA GTCGACGACC GGACCATGGC CACGCCCGCC TACTGGGTCC
 8221 GGCACGTCCG GGAAGCGGTC CGCTTCGGCG ACGGCATCCG GGCACTCGGG AAACTGGGCA
 8281 CCGGCAGCTT CCTGGAAGTC GGGCCGGACG GCGTCCTCAC CGCCATGGCG CGCGCATGCG
 8341 TCACCGCCGC CCCGGAGCCC GGCCACCGCG GCGAACAGGG CGCCGATGCC GACGCCCACA
 8401 CCGCGTTGCT GCTGCCCGCC CTGCGCCGAG GACGGGACGA GGCGCGATCG CTCACCGAGG
 8461 CCGTGGCACG GCTCCACCTG CACGGCGTGC CGATGGACTG GACCTCCGTC CTCGGCGGCG
 8521 ACGTGAGCCG GGTCCCCCTC CCGACGTACG CCTTCCAACG CGAATCCCAC TGGCTGCCGT
 8581 CCGGAGAGGC TCACCCGCGA CCGGCGGACG ACACCGAATC CGGCACGGGA CGGACCGAGG
 8641 CGTCCCCGCC GCGGCCGCAC GACGTCCTGC ACCTCGTGCG CTCCCACGCG GCGGCTGTGC
 8701 TCGGACATTC CCGGGCCGAG CGGATCGACC CCGACCGCGC GTTCCGCGAC CTCGGCTTCG
 8761 ACTCGCTGAC GGCGCTGGAA CTGCGGGACC GGCTCGACAC CGCACTCGGC CTCCGCCTGC
 8821 CCAGCAGCGT GCTCTTCGAC CACCCGAGCC CCGGCGCACT GGCACGCTTC CTCCAGGGCG
 8881 ACGACACGAG GCGCCCCGAA CCAGGGAAGA CGAACGGCAC GCGCGCCACG GAGCCAGGCC
 8941 CGGACCCGGA CGACGAGCCG ATCGCCATCG TCGGCATGGC GTGCCGCTTC CCGGGTGGCG
 9001 TGACCTCTCC GGAGGACCTG TGGCGCCTGC TCGCCGCAGG CGAGGACGCG GTGTCCGGCT
 9061 TCCCCACGGA CCGGGGCTGG AACGTCACTG ACTCCGCCAC GCGCCGCGGA GGCTTCCTGT
 9121 ACGACGCCGG CGAGTTCGAT GCCGCCTTCT TCGGTATCTC GCCGCGTGAG GCGTTGGTGA
 9181 TGGACCCGCA GCAGCGGTTG CTGCTGGAGA CGTCCTGGGA GGCCCTCGAA CGCGCGGGCG
 9241 TGAGCCCCGG CAGTCTGCGC GGCAGCGACA CGGCCGTGTA CATCGGAGCC ACAGCGCAGG
 9301 ACTACGGCCC CCGACTGCAC GAGTCGGACG ACGACTCGGG CGGCTACGTC CTGACCGGCA
 9361 ATACCGCCAG CGTGGCCTCC GGCCGCATCG CCTACTCCCT CGGTCTGGAG GGGCCTGCGG
 9421 TCACGGTGGA CACGGCGTGT TCGTCGTCGC TGGTGGCACT GCACCTGGCG GTGCAGGCGC
 9481 TGCGCCGTGG CGAGTGCTCA CTGGCATTGG CCGGCGGAGC CACGGTGATG CCTTCGCCCG
 9541 GCATGTTCGT GGAGTTCTCA CGGCAAGGGG GCCTCTCCGA GGACGGCCGC TGCAAGGCGT
 9601 TCGCCGCGAC GGCGGACGGC ACCGGCTGGG CCGAGGGTGT GGGTGTGTTG TTGGTGGAGC
 9661 GGTTGTCGGA TGCGCGGCGG TTGGGTCATC GGGTGTTGGC GGTGGTGCGG GGGAGTGCGG
 9721 TCAATCAGGA TGGTGCGTCG AATGGGTTGA CGGCGCCGAA TGGTCCGTCG CAGCAGCGGG
 9781 TGATCCGTGC GGCGTTGGCT GACGCGGGTC TGGTTCCTGC TGATGTGGAT GTGGTGGAGG
 9841 CGCATGGTAC GGGGACGCGG TTGGGTGATC CGATCGAGGC TCAGGCGTTG TTGGCGACGT
 9901 ATGGGCAGGG GCGTGCGGGT GGGCGTCCGG TGGTGTTGGG GTCGGTGAAG TCGAACATCG
 9961 GTCATACGCA GGCGGCGGCT GGTGTGGCTG GTGTGATGAA GATGGTGCTG GCGCTGGGGC
10021 GGGGTGTGGT GCCGAAGACG TTGCATGTGG ATGAGCCGTC TGCGCATGTG GACTGGTCGG
10081 CTGGTGAGGT GGAGTTGGCG GTTGAGGCGG TGCCGTGGTC GCGGGGTGGG CGGGTGCGGC
10141 GGGCTGGTGT GTCGTCGTTC GGGATCAGTG GCACGAATGC GCATGTGATC GTGGAGGAGG
10201 CGCCTGCGGA GCCGGAGCCG GAGCCGGAGC GGGGTCCGGG CTCTGTTGTG GGTGTGGTGC
10261 CGTGGGTGGT GTCCGGGCGG GATGCGGGGG CGTTGCGTGA GCAGGCGGCA CGCTTGGCTG
10321 CGCACGTGTC GGGTGTAAGT GCGGTCGATG TGGGCTGGTC GTTGGTGGCC ACGAGGTCGG
10381 TGTTCGAGCA CCGGGCGGTG ATGGTCGGCA GTGAACTCGA TGCCATGGCG GAGTCGTTGG
10441 CCGGCTTCGC TGCGGGTGGG GTTGTGCCGG GGGTGGTGTC GGGTGTGGCT CCGGCTGAGG
10501 GTCGTCGTGT GGTGTTCGTC TTTCCTGGTC AGGGTTCGCA GTGGGTGGGG ATGGCGGCTG
10561 GGTTGCTGGA TGCGTGCCCG GTGTTCGCGG AGGCGGTGGC GGAGTGCGCT GCGGTGCTGG
10621 ACCCGTTGAC CGGTTGGTCG CTGGTCGAGG TGTTGCGCGG TGGTGGTGAG GCTGTTCTTG
10681 GGCGGGTTGA TGTGGTGCAG CCGGCGTTGT GGGCGGTGAT GGTGTCACTG GCCCGGACCT
10741 GGCGGTATTA CGGTGTGGAG CCTGCTGCGG TTGTGGGGCA TTCGCAGGGT GAGATTGCTG
10801 CGGCTTGTGT GGCTGGGGGG TTGAGTCTGG CCGATGGTGC GCGGGTGGTG GTGTTGCGGA
10861 GCCGGGCGAT CGCCCGGATC GCTGGTGGGG GCGGCATGGT CTCCGTCAGC CTGCCGGCCG
10921 GCCGTGTCCG CACCATGCTG GAGGAGTTCG ACGGCAGGGT TTCCGTTGCG GCGGTCAACG
10981 GTCCGTCCTC GACCGTGGTG TCGGGTGACG TCCAGGCCCT GGATGAGTTG TTGGCCGGTT
11041 GTGAGCGGGA GGGTGTCCGG GCTCGTCGTG TCCCGGTGGA CTATGCCTCC CACTCCGCGC
```

```
11101 AGATGGACCA GTTACGCGAT GATCTGCTGG AAGCGCTGGC GACGATCGTC CCTACATCGG
11161 CGAACGTACC GTTCTTCTCG ACGGTGACGG CGGACTGGCT GGACACGACC GCTCTGGATG
11221 CGGGGTACTG GTTCACGAAT CTGCGGGAGA CGGTCCGGTT CCAAGAAGCC GTCGAAGGGC
11281 TCGTGGCTCA GGGGATGGGC GCGTTCGTCG AGTGCAGCCC GCACCCCGTC CTCGTCCCGG
11341 GCATCACAGA AACACTCGAC ACCTTCGACG CCGACGCTGT CGCACTGTCG TCGCTGCGGC
11401 GTGACGAAGG CGGCCTGGAT CGGTTCCTCA CGTCCCTCGC GGAAGCCTTC GTCCAGGGCG
11461 TCCCGGTCGA CTGGTCCCGC GCCTTCGAGG GTGCGAGCCC CCGCACCGTC GACCTGCCCA
11521 CCTACCCCTT CCAACGGCAA CGCTACTGGC TGCTCGACAA GGCGGCGCAA CGGGAACGCG
11581 AGCGGCTGGA GGACTGGCGC TACCACGTCG AGTGGCGCCC CGTCACGACA CGACCTTCCG
11641 CACGGCTGTC CGGTGTCTGG GCCGTGGCGA TTCCGGCACG TCTGGCCCGT GACTCACTGT
11701 TGGTCGGCGC CATCGACGCA CTGGAGCGAG GCGGCGCCCG TGCCGTGCCC GTGGTGGTCG
11761 ATGAGCGGGA CCACGACCGG CAAGCGCTGG TCGAGGCTCT GCGGAACGGG CTGGGCGACG
11821 ACGACCTCGC CGGTGTGCTC TCCCTTTTGG CCCTCGACGA AGCCCCGCAC GGTGACCACC
11881 CCGACGTGCC CGTCGGCATG GCCGCTTCGC TGGCGCTCGT GCAGGCGATG GCCGACGCCG
11941 CGGCCGAGGT GCCCGTATGG TTCGCGACCC GAGGCGCCGT AGCGGCACTG CCCGGTGAGT
12001 CACCGGAGCG ACCCAGGCAG GCGCTGCTCT GGGGACTGGG ACGGGTCGTC GCCCTGGAAC
12061 AGCCGCAGAT ATGGGGCGGG TTGGTCGACC TCCCGCAACA CCTGGACGAG GACGCGGGCC
12121 GACGGCTGGT CGATGTCGTG GGCGGCCTGG CGGACGAGGA CCAGCTTGCC GTACGGGCCT
12181 CCTCCGTCCT CGCCCGACGC CTCGTTCGTA CGCCGGGTCA CCGTATGTCG AGCCAGGCGG
12241 GCGGGCGCGA GTGGTCGCCC AGCGGCACGG TCCTGGTGAC CGGAGGCACC GGGGCGCTGG
12301 GCGCGCACGT CGCCCGCTGG CTGGCCGGCA AGGGCGCCGA GCACCTGGTA CTCATCAGCC
12361 GTCGCGGAGC GGACGCAGCC GGGGCCGCTG CCCTTCGGGA CAGCCTCACG GACATGGGTG
12421 TCCGGGTGAC CCTGGCCGCG TGCGATGCAG CGGACCGGCA CGCACTGGAG ACGCTCCTCG
12481 ACTCGCTGCG CACGGATCCG GCGCAGCTGA CGGCCGTCAT CCACGCCGCG GGTGCTCTGG
12541 ACGACGGCAT GACGACGGTG CTCACACCGG AGCAGATGAA CAACGCCCTG CGAGCGAAAG
12601 TCACGGCCAC CGTCAACCTG CACGAACTGA CCCGGGACCT CGACCTCTCG GCCTTCGTAC
12661 TGTTCTCGTC CATCTCCGCC ACCCTGGGAA TCCCCGGGCA GGCCAACTAC GCGCCGGGAA
12721 ACTCGTTCTT GGACGCCTTC GCGGAATGGC GCAGGGCTCA GGGGCTCGTG GCGACCTCCA
12781 TCGCCTGGGG ACCGTGGTCC GGCGGCACCG GCATGGCACA TGAAGGGTCG GTGGGCGAAC
12841 GGCTCCAGCG GCACGGTGTA CTCGCCATGG AACCCGCGGC GGCCATCGCT GCGCTCGACC
12901 ACACGCTGGC GAGCGACGAA ACCGCAGTGG CCGTGGCCGA CATCGACTGG AGCCGGTTCT
12961 TCCTGGCGTA CACAGCACTG CGGGCACGGC CCTTGATCGG AGAGATACCC GAGGCACGCC
13021 GCATGCTGGA GTCCGGCTCA GGCCCCGGCG ACCTCGAGCC GGACCGTGCC GAACCCGAGC
13081 TTGCCGTGCG TCTCGCGGGC CTCACCGCGG TCGAGCAGGA ACGTCTTCTG GTGCAGCTCG
13141 TGAGGGAGCA GGCCGCCGTC GTCCTCGGAC ATTCCGGCGC CGAGGCGGTG GCTCCGGACC
13201 GAGCGTTCAA GGATCTCGGA TTCGACTCGC TGACCTCGGT CGAACTGCGC AACCGGCTGA
13261 ACACCGCCAC CGGCCTCAGA CTGCCCGTGA CGGCCGTCTT CGACTACGCG AGGCCCGCGG
13321 CGCTGGCCGG CCATCTGCGC TCCAGGCTGA TCGACGACGA TGGTGACCAC GGTGCCTTGC
13381 CCGGCGTGGA GAAGCACGCG ATCGACGAGC CGATCGCGAT CGTGGGAATG GCATGCCGCT
13441 TCCCGGGAGG CATCGCTTCC CCGGAGGATC TGTGGGACGT GCTCACCGCT GGTGAGGACG
13501 TTGTCTCCGG ACTGCCGCAG AACCGCGGGT GGGACTTGGG GCGCCTGTAC GATCCCGATC
13561 CGGACCGGGC CGGTACGTCA TACATGCGTG AGGGTGCTTT CCTGCACGAG GCGGGGGAGT
13621 TCGACGCGGC CTTCTTCGGT ATCTCGCCGC GTGAGGCGTT GGCGATGGAC CCGCAGCAGC
13681 GGTTGCTGCT GGAGACGTCC TGGGAGGCCC TCGAACGGGC CGGCATCACT CCTTCCAAGC
13741 TGGCGGGCAG TCCGACCGGT GTGTTCTTCG GCATGTCGAA CCAGGACTAC GCCGCCCAGG
13801 CGGGCGACGT GCCGTCCGAG CTGGAGGGCT ACCTGCTCAC CGGCTCCATC TCCAGCGTCG
13861 CTTCGGGGCG TGTTGCTTAC ACGTTCGGTC TTGAGGGGCC TGCGGTGACG GTGGATACGG
13921 CGTGTTCGTC GTCGTTGGTG GCGTTGCATC TGGCGGTGCA GGGGTTGCGG CGGGGTGAGT
13981 GTTCGCTTGC GTTGGTGGGT GGGGTGACGG TGATGTCGTC GCCGGTGACG TTGACGACGT
14041 TCAGTCGGCA GCGGGGTTTG TCGGTGGATG GGCGGTGCAA GGCGTTCGCG GCTTCGGCGG
14101 ATGGTTTTGG TGCTGCCGAG GGTGTGGGTG TGTTGTTGGT GGAGCGGTTG TCGGATGCGC
14161 GGCGGTTGGG TCATCGGGTG TTGGCGGTGG TGCGGGGGAG TGCGGTCAAT CAGGATGGTG
14221 CGTCCAATGG TCTGGCGGCG CCGAATGGTC CGTCGCAGCA GCGGGTGATC CGTGCGGCGT
14281 TGGCTGACGC GGGTCTGGCT CCTGCCGATG TGGATGTGGT GGAGGCGCAT GGCACGGGGA
14341 CGCGGTTGGG TGATCCGATC GAGGCTCAGG CGTTGCTGGC GACGTATGGG CAGGGTCGTA
14401 CCAGTGGGCG TCCGGTGTGG CTGGGGTCGG TGAAGTCGAA CATCGGGCAT ACGCAGGCGG
14461 CGGCCGGTGT GGCTGGTGTG ATGAAGATGG TGCTGGCGTT GGGTCGGGGT GTGGTGCCGA
14521 AGACGTTGCA TGTGGATGAG CCGTCACCGC ATGTGGACTG GTCGGCTGGT GAGGTGGAGT
14581 TGGCGGTTGA GGCGGTGCCG TGGTCGCGGG GTGGGCGGGT GCGGCGGGCT GGTGTGTCGT
14641 CGTTCGGGAT CAGCGGCACG AATGCGCATG TGATCGTGGA GGAGGCGCCT GCGGAGCCTT
14701 CGGTGGAGGA GGGTCCGGGC TCCGTTGTGG GTGTGGTGCC GTGGGTGGTG TCCGGGCGGG
14761 ATGCGGGGGC GTTGCGTGCA CAGGCGGCAC GCTTGGCTGC GCACGTGTCG AGCACGGGTG
14821 CGGGTGTGGT TGATGTGGGC TGGTCGTTGG TGGCCACGAG GTCGGTGTTC GAGCACCGGG
14881 CGGTAATGGT CGGCACTGAT CTTGATTCCA TGGCGGGGTC GTTGGCCGGC TTCGCTGCGG
14941 GTGGTGTTGT GCCGGGGGTG GTGTCGGGTG TGGCTCCGGC TGAGGGCCGT CGTGTGGTGT
15001 TCGTCTTTCC TGGTCAGGGT TCGCAGTGGG TGGGGATGGC GGCTGGGTTG CTGGATGCGT
15061 GTCCGGTGTT CGCGGAGGCG GTGGCGGAGT GTGCCGCGGT GCTGGACCGG TTGACCGGTT
```

```
15121 GGTCGCTGGT CGAGGTGTTG CGTGGTGGTG AGGCTGTTCT TGGGCGGGTT GATGTGGTGC
15181 AGCCGGCGTT GTGGGCGGTG ATGGTGTCAC TGGCTCGGAC CTGGCGGTAT TACGGTGTGG
15241 AGCCTGCTGC GGTTGTGGGG CATTCGCAGG GTGAGATTGC TGCGGCTTGT GTGGCTGGGG
15301 GGTTGAGTCT GGCCGATGGT GCGCGGGTGG TGGTGTTGCG GAGTCGGGCG ATCGCCCGGA
15361 TCGCTGGTGG GGGCGGCATG GTCTCGGTCG GTCTTTCAGC TGAGCGTGTC CGCACCATGC
15421 TCGACACCTA CGGCGGCAGG GTTTCCGTCG CGGCGGTCAA TGGCCCGTCC TCGACCGTGG
15481 TGTCCGGTGA CGCCCAGGCC CTGGATGAGT TGTTGGCCGG TTGTGAGCGG GAGGGTGTCC
15541 GGGCTCGTCG TGTCCCGGTG GACTATGCCT CCCACTCCGC GCAGATGGAC CAGTTACGCG
15601 ATGAGTTGCT GGAGGCGCTG GCGGACGTCA CTCCGCAGGA CTCCAGTGTT CCGTTTTTCT
15661 CGACGGTGAC GGCGGACTGG CTGGACACGA CCGCTCTGGA TGCGGGGTAC TGGTTCACGA
15721 ATCTGCGGGA GACGGTCCGG TTCCAGGAAG CCGTTGAAGG GCTTGTGGCT CAGGGGATGG
15781 GCGCGTTCGT CGAGTGCAGC CCGCACCCTG TCCTCGTCCC GGGCATCACA GAAACACTCG
15841 ACACCTTCGA CGCCGACGCT GTCGCACTGT CGTCGCTGCG GCGTGACGAA GGCGGCCTGG
15901 ATCGGTTCCT CACGTCCCTC GCGGAAGCCT TCGTCCAAGG CGTTCCCGTC GACTGGACCC
15961 ATGCCTTCGA GGGTGGACGC CCGCGCTTCG TCGACCTGCC CACCTATGCC TTCCAGCGAC
16021 AGCGCTACTG GCTGCACGAA GAGCCGCTGC AAGAGCCGGT CGATGAGGCG TGGGATGCCG
16081 AGTTCTGGTC TGTGGTCGAA CGCGGCGATG CCACAGCCGT GTCCGACTTG CTGAGCACGG
16141 ACGCCGAGGC TTTGCACACG GTGTTGCCGG CTTTGTCGTC GTGGCGGCGG CGTCGGGTGG
16201 AGCATCGACG GCTTCAGGAC TGGCGTTACC GGGTGGAGTG GAAGCCTTTC CCGGCCGCGC
16261 TTGATGAGGT GCTCGGTGGT GGCTGGTTGT TCGTGGTGCC GCGGGGCTTG GCGGATGATG
16321 GTGTGGTTGC GCGGGTGGTG GCTGCCGTCA CGGCGCGGGG TGGCGAGGTC AGTGTCGTGG
16381 AGCTCGATCC GACCCGTCCT GACCGCCGGG CTTATGCGGA GGCTGTCGCG GGCCGTGGTG
16441 TGAGCGGGGT CGTGTCGTTC TTGTCCTGGG ATGATCGGCG GCACTCGGAG CATTCTGTTG
16501 TTCCCGCCGG TCTTGCCGCG TCGCTGGTGT TGGCGCAGGC GTTGGTTGAT CTTGGCCGGG
16561 TTGGTGAGGG GCCGCGGTTG TGGCTGGTGA CGCGGGGTGC GGTGGTTGCT GGTCCTTCGG
16621 ATGCCGGTGT GGTGATTGAT CCGGTGCAGG CGCAGGTGTG GGGTTTCGGG CGTGTTCTGG
16681 GTCTGGAGCA TCCCGAGTTG TGGGGTGGGC TGGTGGACCT GCCGGTGGGG GTTGATGAGG
16741 AGGTGTGCCG GCGGTTCGTG GGTGTTGTGG CGTCGGCTGG TTTTGAGGAT CAGGTGGCGG
16801 TGCGTGGTTC GGGTGTGTGG GTGCGTCGTC TGGTGCGTGC TGTGGTGGAT GGTGGTGGGG
16861 GTGGTTGGCG GCCGCGTGGG ACGGTGTTGG TCACGGGTGG TCTTGGTGGT TTGGGTGCGC
16921 ATACGGCCCG GTGGTTGGTG GGTGGTGGGG CGGATCATGT GGTTCTTGTG AGCCGTCGTG
16981 GTGGCAGTGC GCCTGGTGCT GGGGATCTGG TGCGGGAGCT GGAGGGGTTG GGCGGGGCTC
17041 GGGTGTCGGT GCGGGCCTGT GATGTGGCTG ATCGTGTGGC GTTGCGGGCG TTGTTGTCGG
17101 ATCTGGGTGA GCCGGTGACG GCGGTGTTCC ATGCGGCTGG TGTTCCTCAG TCGACGCCTT
17161 TGGCGGAGAT CTCTGTCCAG GAGGCGGCTG ATGTGATGGC GGCCAAGGTG GCGGGTGCGG
17221 TGAATCTGGG TGAGTTGGTG GATCCCTGTG GTCTGGAGGC GTTTGTGTTG TTCTCCTCCA
17281 ATGCCGGTGT GTGGGGCAGT GGGGGGCAGG CGGTGTATGC GGCGGCGAAT GCGTTTCTTG
17341 ATGCGTTGGC GGTGCGTCGT CGGGGTGTTG GTCTGCCGGC CACGAGTGTG GCGTGGGGGA
17401 TGTGGGCTGG TGAGGGGATG GCGTCGGTGG GTGGTGCGGC GCGGGAGTTG TCCCGTCGGG
17461 GGGTGCGGGC GATGGATCCC GAGCGTGCTG TGGCGGTGAT GGCTGATGCG GTGGGTCGTG
17521 GTGAGGCGTT CGTCGCGGTC GCTGATGTGG ACTGGGAACG TTTCGTCACC GGTTTCGCTT
17581 CTGCCCGTCC CCGTCCGTTG ATCAGTGACC TGCCGGAGGT GCGTGCTGTT GTGGAGGGCC
17641 AGGTCCAGGG CCGGGGCCAG GGGTTGGGCT TGGTCGGTGA GGAGGAGTCG TCGGGGTGGT
17701 TGAAGCGGTT GTCGGGGTTG TCTCGTGTGC GGCAGGAGGA GGAGTTGGTG GAGTTGGTCC
17761 GTGCTCAGGC TGCCGTTGTT CTCGGGCATG GTTCCGCGCA GGACGTCCCG GCTGAGCGGG
17821 CGTTCAAGGA GTTGGGTTTT GATTCCCTCA CTGCTGTCGA GCTACGCAAC GGGCTGGCCG
17881 CGGCCACCGG GATCCGGCTG CCGGCCACCA TGGCATTCGA TCATCCCACC GCCACCGCCA
17941 TCGCACGCTT CCTGCAATCC GAACTCGTGG GAAGTGACGA CCCGCTGACG CTCATGCGGT
18001 CGGCGATCGA CCAGTTGGAG ACCGGTCTGG CTCTGCTGGA ATCGGACGAA GAAGCTCGCT
18061 CGGAAATCAC GAAGCGATTG AACATTCTTC TGCCCCGCTT CGGAAGCGGA GGCAGTTCGA
18121 GAGGCAGGGA AGCAGGACAA GACGCAGGCG AACATCAGGA TGTCGAGGAC GCCACCATCG
18181 ATGAGCTATT CGAGGTGCTC GACAACGAAC TCGGCAATTC CTGAAAACCT GTCCGACTGC
18241 TACCGCGACC TTGACCGGAG AACGCTGTGA CGAACGACGA AAAGATCGTC GAGTATCTCA
18301 AGCGCGCGAC CGTGGACCTG CGCAAGGCCC GGCACCGCAT CTGGGAGCTG GAGGACGAGC
18361 CCATCGCGAT CACGTCGATG GCCTGCCACT TCCCGGGCGG GATCGAGAGT CCGGAGCAGC
18421 TGTGGGAACT CCTGTCCGCC GGAGGCGAGG TGCTTTCCGA GTTCCCCGAC GACCGCGGCT
18481 GGGACCTGGA CGAGATCTAC CATCCTGACC CGGAACACAG TGGGACGAGC TACGTCCGTC
18541 ACGGCGGTTT CCTGGATCAT GCGACGCAGT TCGACACGGA CTTCTTCGGT ATCTCGCCGC
18601 GTGAGGCGTT GGCGATGGAC CCGCAGCAGC GGTTGCTGCT GGAGACGTCC TGGCAGCTTT
18661 TCGAGCGCGC AGGAGTCGAT CCCCATACGC TGAAGGGAAG CCGGACCGGA GTATTCGTCG
18721 GCGCCGCACA CATGGGTTAT GCGGACAGGG TGGACACTCC GCCGGCGGAG GCCGAGGGCT
18781 ACCTGCTGAC AGGGAACGCC TCGGCCGTTG TCTCCGGGCG TATTTCCTAC ACCTTCGGCC
18841 TTGAGGGGCC TGCGGTGACG GTGGACACGG CGTGCTCGTC GTCGCTGGTG GCGCTGCACC
18901 TGGCGGTGCA GGCGCTGCGC CGTGGCGAGT GCTCGCTGGC GGTCGTCGGT GGTGTGGCCG
18961 TCATGTCGGA CCCGAAGGTC TTCGTCGAGT TCAGCCGGCA GCGCGGACTG GCCAGGGACG
19021 GCCGGTCCAA GGCTTTTGCG GCGTCAGCGG ATGGTTTCGG CTTCGCCGAG GGAGTTTCGC
19081 TGCTCTTGCT GGAGCGGTTG TCGGATGCGC GGCGGTTGGG TCATCGGGTG TTGGCGGTGG
```

```
19141 TGCGGGGGAG TGCGGTCAAT CAGGATGGTG CGTCCAATGG TCTGGCGGCG CCGAATGGTC
19201 CGTCGCAGCA GCGGGTGATT CGTGCGGCGT TGGCTGACGC GGGTCTGGCT CCTGCCGATG
19261 TGGATGTGGT GGAGGCGCAT GGTACGGGGA CGCGGTTGGG TGATCCGATC GAGGCTCAGG
19321 CGTTGCTGGC GACGTATGGG CAGGGGCGTA CCAGTGGGCG TCCGGTGTGG CTGGGGTCGG
19381 TGAAGTCGAA CATCGGTCAT ACGCAGGCGG CGGCCGGTGT GGCTGGTGTG ATGAAGATGG
19441 TGCTGGCTCT GGAGCGGGGT GTGGTGCCGA AGACGTTGCA CGTGGATGAG CCGTCTCCGC
19501 ATGTGGACTG GTCGACCGGT GCGGTGGAGT TGCTGACTGA AGAGCGGCCG TGGGAGCCGG
19561 AGGCTGAGCG TCTTCGTCGG GCAGGCATTT CCGCCTTCGG TGTCAGTGGC ACGAATGCGC
19621 ATGTGATCGT GGAGGAGGCA CCTGCGGAAC CGGAACCGGA GCCGGAGCCG GGAACTCGTG
19681 TGGTTGCTGC CGGTGATCTG GTGGTGCCGT GGGTGGTGTC CGGGCGGGAT GCGGGGGCGT
19741 TGCGTGCACA GGCGGCACGC TTGGCTGCGC ATGTGTCGAG CACGGGTGCG GGTGTGGTTG
19801 ATGTGGGCTG GTCGTTGGTG GCCACGAGGT CGGTGTTCGA GCACCGGGCG GTGATGGTCG
19861 GCACTGATCT TGATTCCATG GCGGGGTCGT TGGCCGGGTT TGCTGCGGGT GGGGTTGTGC
19921 CGGGGGTGGT GTCGGGTGTG GCTCCGGCTG AGGGTCGTCG TGTGGTGTTC GTCTTTCCTG
19981 GTCAGGGTTC GCAGTGGGTG GGGATGGCGG CTGGGTTGCT GGATGCGTGT CCGGTGTTCG
20041 CGGAGGCGGT GGCGGAGTGT GCCGCGGTGC TGGACCCGTT GACCGGTTGG TCGCTGGTCG
20101 AGGTGTTGCG CGGTGGTGAG GCTGTTCTTG GGCGGGTTGA TGTGGTGCAG CCGGCGTTGT
20161 GGGCGGTGAT GGTGTCACTG GCTCGGACCT GGCGGTATTA CGGTGTGGAG CCTGCTGCGG
20221 TTGTGGGGCA TTCGCAGGGT GAGATTGCTG CGGCTTGTGT GGCTGGGGGG TTGAGTCTGG
20281 CCGATGGTGC GCGGGTGGTG GTGTTGCGGA GCCGGGCGAT CGCCCGGATC GCCGGTGGGG
20341 GCGGCATGGT CTCCGTCAGT CTCCCGGCCG GCCGTGTCCG CACCATGCTC GACACCTACG
20401 GCGGCCGGTT GTCGGTGGCT GCGGTCAACG GCCCGTCCTC GACCGTGGTG TCCGGTGACG
20461 CCCAGGCCCT GGATGAGTTG TTGGCCGGCT GTGAGCGGGA GGGGGTCCGG GCTCGTCGTG
20521 TCCCGGTGGA CTATGCCTCC CACTCCGCGC AGATGGACCA GTTACGCGAT GAGCTGCTGG
20581 AAGCGCTGGC GGACATCACT CCGCAACACT CCAGCGTTCC GTTCTTCTCG ACGGTGACGG
20641 CGGACTGGCT GGACACGACC GCTCTGGATG CGGGGTACTG GTTCACGAAT CTGCGGGAGA
20701 CGGTCCGGTT CCAGGAAGCC GTCGAAGGGC TTGTGGCTCA GGGGATGGGC GCGTTCGTCG
20761 AGTGCAGCCC ACACCCCGTC CTCGTCCCCG GTATCGAGCA GACCCTCGAC ACCGTGGAAG
20821 CCGATGCTGT GGCGCTGGGT TCGCTACGGC GTGATGAGGG CGGCCTGGGA CGGTTCCTCA
20881 CGTCCCTCGC GGAAGCCTTC GTCCAGGGCG TCCCGGTCGA CTGGTCCCGC ACCTTCGAGG
20941 GTGCGAGCCC CCGCACCGTC GACCTGCCCA CCTATCCCTT CCAACGGCAA CGTTTCTGGT
21001 TGGAGGGATC CCCGGCGTTG TCTTCGAACG GCGTCGAGGG TGAGGCGGAC GTCGCGTTCT
21061 GGGATGCGGT CGAGCGCGAG GACTCGGCGG TTGTAGCCGA GGAGTTGGGG ATCGACGCCA
21121 AGGCTCTGCA CATGACATTG CCGGCCTTGT CGTCGTGGCG GCGGCGTGAG CGGCAGCGTC
21181 GGAAGGTGCA GCGCTGGCGT TACCGGGTGG AGTGGAAGCG TCTCCCGAAT TCGCGGGCAC
21241 AGGAGTCGCT GCAGGGCGGC TGGTTGCTCG TCGTCCCGCA GGGCCGTGCC GGCGATGTCC
21301 GCGTCACTCA GTCGGTGGCG GAGGTGGCGG CCAAGGGTGG TGAAGCCACG GTCCTGGAGG
21361 TCGACGCCCT GCATCCCGAC CGCGCAGCAT ACGCCGAGGC CCTCACCCGG TGGCCGGGTG
21421 TGCGGGGTGT GGTGTCGTTC CTGGCGTGGG AGGAGCAGGC CCTTGCCGAA CACCCCGTTC
21481 TGTCTGCGGG TCTGGCGGCA TCGCTGGCGT TGGCCCAGGC GTTGATCGAT GTCGGCGGGT
21541 CCGGTGAGTC GGCGCCGCGT CTGTGGCTGG TCACGGAAGC TGCCGTCGTG ATCGGTGCTG
21601 CCGACACCGG TGCGGTGATC GACCCCGTAC ACGCGCAGCT GTGGGGCTTC GGCCGTGTCC
21661 TTGCTCTGGA ACACCCCGAA TTGTGGGGCG GGCTGATCGA CCTGCCCGCT GTGGCAGGCG
21721 AGCCTGGTTC GATTACCGAC CACGCGCATG CCGACCTACT GGCCACGGTC CTGGCCACGA
21781 TGGTGCAGGC TGCTGCCCGA GGCGAGGACC AGGTCGCGGT CCGGACGACC GGTACTTACG
21841 TACCCAGGCT GGTGCGTTCA GGCGGCAGTG CACACTCGGG TGCGCGGAGG TGGCAGCCGC
21901 GCGACACCGT ACTGGTCACC GGCGGGATGG GACCGCTGAC CGCCCACATC GTCCGTTGGC
21961 TGGCTGACAA CGGTGCCGAC CAGGTAGTAC TCCTGGGAGG TCAGGGAGCA GACGGCGAGG
22021 CCGAGGCGCT GAGGGCCGAG TTCGACGGGC ACACGACGAA GATCGAACTC GCGGACGTGG
22081 ACACCGAGGA CAGCGACGCG CTGCGGTCCT TGCTCGACCG CACGACCGGC GAACACCCGC
22141 TGCGCGCGGT CATCCATGCG CCGACCGTGG TCGAGTTCGC CTCGGTGGCC GAGTCGGACC
22201 TGGTGCGATT CGCCCGCACC ATCAGCAGCA AGATCGCCGG CGTCGAGCAG CTCGACGAGG
22261 TGCTGAGCGG CATCGACACG GCGCACGACG TGGTCTTCTT CTCCTCCGTC GCGGGCGTCT
22321 GGGGAAGCGC GGGGCAGAGC GCCTACGCGG CGGGCAACGC CTTCCTCGAC GCCGTCGCCC
22381 AGCACCGCCG TCTGCGCGGA CTGCCCGGTA CGTCGGTGGC CTGGACTCCG TGGGACGACG
22441 ATCGATCCCT TGCCTCCCTC GGTGACTCGT ACCTCGACCG ACGAGGACTG CGAGCACTGT
22501 CCATACCCGG CGCGCTCGCC TCCCTCCAGG AAGTGCTCGA CCAGGACGAG GTCCACGCCG
22561 TGGTGGCGGA TGTCGACTGG GAGCGGTTCT ACGCCGGCTT CAGTGCCGTC CGGCGCACTT
22621 CCTTCTTCGA CGACGTGCAC GACGCCCACC GGCCGGCCCT GTCCACGGCT GCGACCAACG
22681 ACGGACAGGC CCGGGACGAG GACGGCGGTA CGGAACTCGT ACGACGTCTG CGTCCGCTGA
22741 CCGAGACGGA GCAACAGCGA GAGCTCGTGT CGCTCGTCCA GAGTGAAGTC GCTGCCGTCC
22801 TAGGCCACTC CTCCACCGAC GCGGTCCAGC CACAGCGCGC GTTCCGAGAG ATCGGGTTCG
22861 ACTCACTGAC AGCGGTCCAG CTCCGGAACC GGCTTACGGC CACCACGGGC ATGCGCCTTC
22921 CGACAACGCT GGTCTTCGAC TACCCGACCA CCAACGGACT CGCCGAGTAC CTGCGCTCCG
22981 AACTGTTCGG TGTGTCCGGC GCACCAGCTG ACCTCTCCGT CGTCCGGAAC GCGGATGAGG
23041 AGGACGACCC CGTCGTCATC GTGGGGATGG CCTGCCGGTT CCCGGGCGGG ATCGATACGC
23101 CGGAAGCCTT CTGGAAGCTG CTCGAAGCGG GCGGCGATGT CATCTCCGAA CTTCCGGCCA
```

```
23161 ACCGCGGCTG GGACATGGAG CGACTCCTGA ACCCGGACCC CGAGGCGAAG GGCACCAGCG
23221 CCACACGCTA CGGCGGTTTC CTCTACGACG CCGGGGAGTT CGACGCCGCC TTCTTCGGTA
23281 TCTCGCCGCG TGAGGCGTTG GCGATGGACC CGCAGCAACG GCTGCTGCTG GAAACCGTCT
23341 GGGAGCTCAT CGAGAGCGCC GGCGTGGCGC CCGACTCGCT CCACCGGAGC CGGACCGGCA
23401 CGTTCATCGG CAGCAACGGC CAGTTCTACG CACCGCTGCT GTGGAACTCC GGCGGTGATC
23461 TGGAGGGCTA CCAAGGCGTG GGCAACGCCG GCAGCGTCAT GTCCGGCCGC GTCGCCTACT
23521 CCCTCGGTCT TGAGGGGCCT GCGGTGACGG TGGATACGGC GTGTTCGTCG TCGCTGGTGG
23581 CACTGCACCT GGCGGTGCAG GCGCTGCGCC GTGGCGAGTG CTCACTCGCC ATAGCCGGCG
23641 GTGTGACGGT GATGTCCACA CCGGACAGCT TCGTTGAGTT CTCACGGCAA CAGGGCCTTT
23701 CCGAGGACGG CCGTTGCAAG GCGTTCGCGA GCACAGCCGA TGGTTTCGGC CTCGCCGAGG
23761 GCGTTTCGGC GCTGTTGGTG GAGCGGTTGT CGGATGCGCG GCGGTTGGGT CATCGGGTGT
23821 TGGCGGTGGT GCGGGGGAGT GCGGTCAATC AGGATGGTGC GTCGAATGGG TTGACGGCGC
23881 CGAATGGTCC GTCGCAGCAG CGGGTGATTC GTGCGGCGTT GGCTGACGCG GGTCTGGCTC
23941 CTGCTGATGT GGATGTGGTG GAGGCGCATG GTACGGGGAC GCGGTTGGGT GATCCGATCG
24001 AGGCTCAGGC GTTGTTGGCG ACGTATGGGC AGGGTCGTGC GGGTGGGCGT CCGGTGGTGT
24061 TGGGGTCGGT GAAGTCGAAC ATCGGGCATA CGCAGGCGGC GGCTGGCGTG GCTGGTGTGA
24121 TGAAGATGGT GCTGGCGCTG GAGCGGGGTG TGGTGCCGAA GACGTTGCAT GTGGATGAGC
24181 CGTCACCGCA TGTGGACTGG TCGGCTGGTG AGGTGGAGTT GGCGGTTGAG GCGGTGCCGT
24241 GGTCGCGGGG TGGGCGGGTG CGGCGGGCTG GTGTGTCGTC GTTCGGGATC AGTGGCACGA
24301 ATGCGCATGT GATTGTGGAG GAGGCGCCTG CGGAGCCGGA GCCGGAGCCG GGAACTCGTG
24361 TGGTTGCTGC TGGTGATCTG GTGGTGCCGT GGGTGGTGTC CGGGCGGGAT GCGGGGGCGT
24421 TGCGTGAGCA GGCGGCCCGG TTGGCTGCGC ACGTGTCGAG CACGGGTGCG GGTGTGGTTG
24481 ATGTGGGGTG GTCGTTGGTG GCCACGAGGT CGGTGTTCGA GCACCGGGCG GTGATGGTCG
24541 GCAGTGAACT CGATTCCATG GCGGAGTCGT TGGCTGGCTT CGCTGCGGGT GGGGTTGTGC
24601 CGGGGGTGGT GTCGGGTGTG GCTCCGGCTG AGGGTCGTCG TGTGGTGTTC GTCTTTCCTG
24661 GTCAGGGTTC GCAGTGGGTG GGGATGGCGG CTGGGTTGCT GGATGCGTGT CCGGTGTTCG
24721 CGGAGGCGGT GGCGGAGTGT GCCGCGGTGC TGGATCCGGT GACGGGTTGG TCGCTGGTCG
24781 AGGTGTTGCG CGGTGGTGGT GAGGCTGTTC TTGGGCGGGT TGATGTGGTG CAGCCGGCGT
24841 TGTGGGCGGT GATGGTGTCA CTGGCCCGGA CCTGGCGGTA TTACGGTGTG GAGCCTGCTG
24901 CGGTTGTGGG GCATTCGCAG GGTGAGATCG CTGCGGCTTG TGTGGCTGGG GGGTTGAGTC
24961 TGGCCGATGG TGCGCGGGTG GTGGTGTTGC GGAGCCGGGC GATCGCCCGG ATCGCTGGTG
25021 GGGGCGGCAT GGTCTCGGTC GGTCTTTCAG CTGAGCGTGT CCGCACCATG CTCGACACCT
25081 ACGGTGGCCG GGTTTCGGTC GCGGCGGTCA ATGGCCCGTC CTCGACCGTC GTGTCCGGTG
25141 ACGTCCAGGC CCTGGATGAG TTGTTGGCCG GTTGTGAGCG GGAGGGTGTC CGGGCTCGTC
25201 GTGTCCCGGT GGACTATGCC TCCCACTCCG CGCAGATGGA CCAGTTACGC GATGAGCTGC
25261 TGGAAGCGCT GGCGGACATC ACTCCGCAAC ATTCCAGTGT TCCGTTCTTC TCGACGGTGA
25321 CGGCGGACTG GCTGGACACG ACCGCTCTGG ATGCGGGGTA CTGGTTCACG AATCTGCGGG
25381 AGACGGTCCG GTTCCAGGAA GCCGTCGAAG GGCTCGTGGC TCAGGGGATG GGCGCGTTCG
25441 TCGAGTGCAG CCCGCACCCC GTCCTCGTCC CCGGTATCGA GCAGACCCTC GACGCCCTCG
25501 ACCAGAACGC CGCCGTACTC GGCTCCCTGC GGCGTGACGA AGGCGGCCTG GACCGACTCC
25561 TCACATCCCT CGCGGAAGCC TTCGTCCAAG GCGTTCCCGT CGACTGGACC CACGCCTTCG
25621 AAGGCATGAC CCCCCGCACC GTCGACCTGC CCACCTACCC CTTCCAACGA CAGCACTACT
25681 GGCCCAAGCC CGCACCGGCC CCCGGCGCGA ACCTGGGCGA CGTGGCGTCC GTGGGCCTCA
25741 CCGCGGCCGG CCACCCCCTT CTGGGCGCGG TCGTGGAGAT GCCCGACTCC GACGGGTTGG
25801 TGCTCACCGG GCAGATCTCC CTGCGGACCC ATCCCTGGCT CGCCGACCAC GAGGTGCTCG
25861 GATCGGTGCT CCTGCCGGGC ACCGCGTTCG TCGAGCTTGC CGTCCAGGCC GCCGACCGCG
25921 CCGGTTACGA CGTACTGGAC GAGCTGACGC TGGAGGCGCC CCTCGTGCTC CCCGACAGGG
25981 GCGGCATCCA GGTGCGTCTG GCCCTCGGGC CGTCCGAGGC AGACGGACGC CGGTCCCTCC
26041 AGCTGCACAG CAGGCCGGAG GAGGCTGCCG GGTTCCACCG CTGGACGAGG CACGCGAGTG
26101 GATTCGTCGT TCCCGGCGGT ACCGGGGCGG CGCGGCCCAC CGAGCCGGCC GGCGTGTGGC
26161 CGCCCGCAGG TGCCGAGCCG GTCGCTCTCG CATCGGACCG GTACGCCCGG CTCGTCGAGC
26221 GCGGCTACAC CTACGGCCCC TCCTTCCAGG GGCTGCACAC CGCATGGCGC CACGGGGACG
26281 ACGTGTACGC GGAAGTGGCG CTGCCAGAAG GAACACCGGC CGACGGCTAC GCCCTGCATC
26341 CGGCCCTGCT GGACGCGGCG GTCCAGGCCG TCGGACTCGG CTCGTTCGTC GAGGATCCCG
26401 GCCAGGTGTA CCTGCCGTTC CTCTGGAGCG ACGTGACGCT GCACGCGACC GGGGCCACGT
26461 CCCTGCGGGT GAGGGTTTCA CCGGCCGGTC CCGACACCGT TGCGCTGGCC CTCGCCGACC
26521 CGGCCGGGGC GCCGGTGGCC ACGGTGGGCG CCCTCCGTCT GCGTACGACG TCCGCGGCGC
26581 AGCTCGCCCG TGCGCGCGGG AGCGCGGAAC ACGCGATGTT CCGCGTGGAG TGGGTGGAGG
26641 AGGGCTCGGC CGCGGACCGG TGCCGGGGCG GCGCGGGCGG GACGACGTAC GAGGGGGAAC
26701 GCGCCGCCGA GGCCGGGGCC GCCGCTGGTA CCTGGGCCGT ACTCGGCCCC CGGGTGCCGG
26761 CCGCCGTCCG GACGATGGGC GTGGATGTCG TCACCGCCCT CGACACGCCG GACCACCCCG
26821 CGGACCCGCA GAGCCTCGCG GACCTGGCGG CGCTCGGGGA CACCGTTCCC GACGTGGTCG
26881 TCGTGACCAG CCTCCTGAGC CTCGCCTCCG GAGCGGATTC CCCCCTAGGG AACCGGCCCC
26941 GGCCGACCGC CGCCGAGCAG GACACCGCCG CCACGGTCGC CGGCGTCCAC AGCGCACTCC
27001 ACGCGGCCCT GGACCTGGTG CAGGCATGGC TGGCCGACGA ACGCCACACC GCCTCCCGGC
27061 TGGTGCTCGT CACCCGGCAC GCGATGACCG TCGCCGAGTC CGACCCCGAG CCTGACCTGC
27121 TCCTCGCCCC GGTGTGGGGA CTCGTGCGGT CCGCCCAGGC CGAGAACCCC GGCCGCTTCG
```

```
27181 TGCTCGCCGA CATCGACGGC GACGAGGCAT CCTGGGATGC TCTGCCCCGA GCCGTCGCCT
27241 CGGCCGCATC GGAGGTGGCG ATACGGGCCG GCGCCGTGTA CGTACCGCGG CTGGCCCGCG
27301 CCACGGACGA GGGACTGGTC GTGGCCGACG AGGCTGCGGG GCCCTGGCGG CTGGACGTCA
27361 CGGAAGCGGG CACCCTGGCG AACCTCGCCC TGGTGCCGTG CCCGGACGCC TCCCGCCCGC
27421 TGGGCCCCGA CGAGGTACGG ATCGCCGTCC GTGCCGCCGG GGTCAACTTC CGGGACGTCC
27481 TCCTGGCCCT GGGCATGTAC CCGGACGAGG GGCTCATGGG CGCGGAGGCG GCGGGCGTCG
27541 TCACCGAGGT CGGCGGGGGC GTCACGACGC TCGCGCCAGG TGACCGGGTG ATGGGCCTGG
27601 TGACCGGTGG ATTCGGGCCG GTGGCCGTGA CGCACCACCG GATGCTCGTA CGGATGCCGC
27661 GTGGCTGGTC CTTCGCCGAG GCCGCGTCGG TGCCGGTGGC GTTCCTGACC GCGTACTACG
27721 CCCTGCACGA CCTGGCAGGC CTGCGCGGCG GCGAGTCGGT GCTGGTGCAC TCCGCTGCGG
27781 GCGGTGTCGG CATGGCGGCC GTGCAGTTGG CACGGCACTG GGATGCCGAG GTGTTCGGCA
27841 CCGCGAGCAA GGGCAAGTGG GACGTTCTCG CGGCGCAGGG CCTCGACGAG GAGCACATCG
27901 GCTCGTCCAG GACGACCGAG TTCGAGCAGC GCTTCCGCGC GACCAGTGGT GGGCGCGGGA
27961 TCGATGTCGT CCTGAATGCC CTCTCGGGTG ACTTCGTCGA CGCCTCGGCG CGTCTCCTGC
28021 GCGAGGGCGG CCGGTTCGTC GAGATGGGCA AGACCGACAT CCGTACCGAC CTCGGCGTCG
28081 TCGGGGCGGA CGGCGTCCCG GACATCCGGT ACGTCGCCTT CGACCTCGCC GAGGCGGGTG
28141 CCGAGCGGAT CGGGCAGATG CTCGACGAGA TCATGGCGCT CTTCGACGCC GGTGTCCTGC
28201 GGTTGCCGCC GTTGCGCGCC TGGCCGGTGC GGCGCGCCCA CGAGGCACTG AGGTTCGTCA
28261 GCCAGGCACG TCATGTGGGC AAGGTCGTCC TCACCGTCCC GGCCGCGCTC GACGCCGAGG
28321 GAACCGTGCT GATCACCGGG GCGGGCACGC TGGGAGCCCT GGTCGCCCGC CACCTCGTCA
28381 CCGAGCACGA CGTCCGCCGG CTGCTGCTGG TCAGCCGCAG CGGCGTCGCC CCCGACCTGG
28441 CGGCCGAACT CGGTGCGCTG GGCGCCGAGG TCACGGTGGC GGCCTGCGAC GTCGCCAACC
28501 GCAAGGCGCT CAAGGCCCTC CTGGAGGACA TACCGCCCGA GCATCCGGTC ACGGGCATCG
28561 TTCACACGGC CGGCGTGCTC GACGACGGTG TGGTGTCCGG GCTCACCCCT GAACGGGTGG
28621 ACACCGTCCT CAAACCCAAG GTGGACGCGG CCCTGACCCT GGAGTCAGTG ATCGGCGAAC
28681 TGGACCTCGA CCCGGCCCTG TTCGTGATCT TCTCATCGGC AGCGAGCATG CTGGGCGGGC
28741 CCGGCCAGGG CAGTTACGCC GCGGCCAATC AGTTCCTGGA CACCCTCGCC CGACACCGGG
28801 CGCGCCGCGG GCTCACCTCC GTGTCACTCG GCTGGGGGCT GTGGCACGAG GCCAGCGGTC
28861 TCACCGGCGG CCTGGCCGAC ATCGACCGTG ACCGGATGAG CCGGGCGGGG ATCGCGCCCA
28921 TGCCGACCGA CGAGGCCCTG CACCTGTTCG ACAGGGCAAC GGAACTCGGC GATCCGGTAC
28981 TCCTGCCGAT GCGCCTGAAC GAGGCCGCGC TGGAGGACCG GGCCGCGGAC GGAACACTGC
29041 CGCCGCTGCT GAGTGGTCTG GTCCGGGTGC GGCACAGGCC GTCGGCGCGG GCAGGTACCG
29101 CGACCGCCGC CCCCGCCACC GGCCCCGAGG CGTTCGCCCG GGAGCTGGCG GCGGCACCGG
29161 ACCCACGTCG TGCCCTGCGC GACCTCGTCC GCGGCCACGT CGCCCTGGTG CTCGGACACA
29221 GTGGCCCCGA GGCCATCGAC GCCGAACAGG CCTTCCGGGA CATCGGTTTC GACTCCCTGA
29281 CCGCAGTCGA ACTCAGAAAC CGGCTGAACG CCGAGACCGG CCTCCGCTTG CCCGGCACGC
29341 TCGTGTTCGA CTACCCCAAC CCGAGCGCGC TCGCCGATCA CCTGCTCGAA CTCCTCGCTC
29401 CCGCGACACA ACCCACCGCA GCCCCGCTGC TCGCCGAACT GGAACGGGTG GAACAACTCC
29461 TGTCTGCGGC CGCGTCACCC GGCGGACCGG CATCCGCGGT GGACGAGGAG ACGCGCACGC
29521 TCATCGCCAC ACGGCTGGCC ACCCTTGCCT CGCAGTGGAC ACACCTCCCG GTCGGTTCGC
29581 CGGGCAACGC GGACAACCGC AGCGGCCCCG GCGAGTCCGG GCAGGCCCAG GAATCCGGAG
29641 CAACCGGGGA GCACACGGCG GCGTGGACGT CGGACGACGA TCTCTTCGCC TTCCTCGACA
29701 AGCGGTTGGA GACGTGATGG CCGCCGGCCG AGTCAGCGAG TCCTTTCGTC CTTCTGCTGG
29761 GGAAAACGAC GCACCGGGAG GTTTTGGTGG CTGAGGCGGA GAAGCTGCGC GAATACCTGT
29821 GGCGCGCCAC GACCGAACTC AAGGAGGTCA GCGATCGACT CCGCGAGACC GAGGAACGGG
29881 CCCGAGAGCC GATCGCCATC GTGGGAATGA GCTGCCGGTT CCCCGGCGGC GGCGACGCCA
29941 CCGTCAACAC GCCCGAACAG TTCTGGGACC TGCTGAACAG CGGCGGTGAC GGCATCGCGG
30001 GTCTACCCGA GGACCGCGGG TGGGACTTGG GGCGCCTGTA CGATCCCGAT CCGGACCGGG
30061 CCGGTACGTC GTACGTGCGT GAGGGCGGTT TCCTGTACGA CTCGGGGGAG TTCGACGCCG
30121 CCTTCTTCGG GATCTCGCCG CGTGAGGCGT TGGCGATGGA CCCGCAGCAG CGGTTGCTGC
30181 TGGAGACGTC CTGGGAGGCA TTCGAGAGCG CCGGTATCAA GCGCGCCGCT CTGAGAGGCA
30241 GCGACACCGG CGTGTACATC GGCGCGTGGA GCACCGGCTA TGCCGGCAGC CCCTACCGCC
30301 TGGTCGAAGG CCTGGAAGGC CAGCTCGCCA TCGGCACCAC ACTAGGGGCC GCTTCGGGGC
30361 GTGTTGCTTA CACGTTCGGT CTTGAGGGGC CTGCGGTGAC GGTGGATACG GCGTGTTCGT
30421 CGTCGTTGGT GGCGTTGCAT CTGGCGGTGC AGGGGTTGCG GCGGGGTGAG TGTTCGCTGG
30481 CGTTGGTGGG TGGGGTGACG GTGATGTCGT CGCCGGTGAC GTTGACGACG TTCAGTCGGC
30541 AGCGGGGTTT GTCGGTGGAT GGGCGGTGCA AGGCGTTCCC GGCTTCGGCG GATGGTTTTG
30601 GTGCTGCCGA GGGTGTGGGT GTGTTGTTGG TGGAGCGGTT GTCGGATGCG CGGCGGTTGG
30661 GTCATCGGGT GTTGGCGGTG GTGCGGGGGA GTGCGGTCAA TCAGGATGGT GCGTCGAATG
30721 GGTTGACGGC GCCGAATGGT CCGTCGCAGC AGCGGGTGAT CCGTGCGGCG TTGGCTGACG
30781 CGGGTCTGGC TCCTGCTGAT GTGGATGTGG TGGAGGCGCA TGGTACGGGG ACGCGGTTGG
30841 GTGATCCGAT CGAGGCTCAG GCGTTGTTGG CGACGTATGG GCAGGGGCGT GCGGGTGGGC
30901 GTCCGGTGTG GCTGGGGTCG GTGAAGTCGA ACATCGGGCA TACGCAGGCG GCGGCCGGTG
30961 TGGCTGGTGT GATGAAGATG GTGCTGGCGC TGGGGCGGGG TGTGGTGCCG AAGACGTTGC
31021 ATGTGGATGA GCCGTCACCG CACGTGGACT GGTCGGCCGG TGCGGTGGAG TTGCTGACTG
31081 AAGAGCGGCC GTGGGAGCCG GAGGCTGAGC GTCTTCGTCG GGCAGGCATC TCCGCCTTCG
31141 GTGTCAGTGG CACGAACGCG CATGTGATCG TGGAGGAGGC GCCTGCGGAA CCGGAGCCGG
```

```
31201 AGCCGGGAAC TCGTGTGGTT GCTGCCGGTG ATCTGGTGGT GCCGTGGGTG GTGTCCGGGC
31261 GGGATGCGAG GGCGTTGCGT GCACAGGCGG CACGCTTGGC TGCGCACGTG TCGGGTGTAA
31321 GTGCGGTCGA TGTGGGCTGG TCATTGGTGG CCACGAGGTC GGTGTTCGAG CACCGGGCTG
31381 TTGCGATCGG CAGTGAACTC GACTCCATGG CGGGTTCGTT GGCCGGCTTC GCTGCGGGTG
31441 GGGTGGTGCC GGGGGTGGTG TCGGGTGTGG CTCCGGCTGA GGGTCGTCGT GTGGTGTTCG
31501 TCTTTCCTGG TCAGGGTTCG CAGTGGGTGG GGATGGCGGC TGGGTTGCTG GATGCGTGTC
31561 CGGTGTTCGC GGAGGCGGTG GCGGAGTGCG CTGCGGTGCT GGATCCGGTG ACGGGTTGGT
31621 CGCTGGTCGA GGTGTTGCAG GGCAGGGACG CGACTGTTCT TGGGCGGGTT GATGTGGTGC
31681 AGCCGGCGTT GTGGGCGGTG ATGGTGTCAC TGGCTCGGAC CTGGCGGTAT TACGGTGTGG
31741 AGCCTGCTGC GGTTGTGGGG CATTCGCAGG GTGAGATTGC TGCGGCTTGT GTGGCTGGGG
31801 GGTTGAGTCT GGCCGATGGT GCGCGGGTGG TGGTGTTGCG GAGCCGGGCG ATCGCCCGGA
31861 TCGCTGGTGG GGGCGGCATG GTCTCCGTCA GCCTGCCGGC CGGCCGTGTC CGCACCATGC
31921 TGGAGGAGTT CGACGGCCGG TTGTCGGTGG CTGCGGTCAA TGGCCCGTCC TCGACCGTGG
31981 TGTCCGGTGA CGTCCAGGCC CTGGATGAGT TGTTGGCCGG TTGTGAGCGG GAGGGTGTCC
32041 GGGCTCGTCG TGTCCCGGTG GACTATGCTT CCCACTCCGC GCAGATGGAC CAGTTACGCG
32101 ATGAGCTGCT GGAGGCGCTG GCGGACATCA CTCCGCAGGA CTCCAGTGTT CCGTTTTTCT
32161 CGACGGTGAC GGCGGACTGG CTGGGCACGA CTGCCCTGGG TGCGGGGTAC TGGTTCACGA
32221 ATCTGCGGGA GACGGTCCGG TTCCAGGAAG CCGTCGAAGG GCTTGTGGCT CAGGGGATGG
32281 GCGCGTTCGT CGAGTGCAGC CCGCACCCCG TCCTCGTCCC CGGTATCGAG CAGACCCTCG
32341 ACGCCCTCGA CCAGAATGCC GCCGTATTCG GCTCGCTGCG GCGTGACGAA GGCGGCCTGG
32401 ACCGGTTTCT CACGTCCCTC GCGGAAGCCT TCGTCCAGGG CGTTCCCGTC GACTGGTCCC
32461 GCGCCTTCGA AGGCGTGACC CCTCGCACCG TCGACCTGCC CACCTACCCC TTCCAACGAC
32521 AGCACTACTG GTTGATGGCG GAAGAGGCAC CGGTCTCTCA GCCCCCTCAC TCGGAGAACA
32581 GCTTCTGGTC GGTAGTGGCC GATGCGGATG CCGAGGCTGC TGCTGAACTT CTGGGTGTCG
32641 ATGTAGAGGC AGTCGAGGCT GTAATGCCGG CGTTGTCTTC GTGGCACCGG CAGAGCCAAC
32701 TTCGTGCCGA AGTCAACCAG TGGCGCTACG ACGTTGCGTG GAAGCGTCTG ACCACCGGGG
32761 CGCTGCCCGA AAAGCCGGGC AACTGGCTCG TCGTGACTCC AGCAGGAACC GACACCACGT
32821 TCGCTGAGTC GTTGGCGAGG ACGGCAGCCG CAGAACTGGG CGTATCCGTC AGCTTTGCGC
32881 AGGTGGACAC TGCTCATCCT GACCGGTCGC AATACGCGCA TGCGCTGCGT CAAGCCCTGA
32941 CCGGCCCGGA GAACGTCGAT CACCTCGTGT CCTTGCTGGC CCTGGACCAG GCCACTGACG
33001 ACCTCGCCGC CGCACCTTCC TGTCTTGCCG CGTCGCTGGT GTTGGCGCAG GCGTTGGTTG
33061 ATCTTGGCCG GGTTGGTGAG GGGCCGCGGT TGTGGCTGGT GACGCGGGGT GCGGTGGTTG
33121 CTGGTCCTTC GGATGCCGGT GCGGTGATTG ATCCGGTACA GGCGCAGGTG TGGGGTTTCG
33181 GGCGTGTTCT GGGTCTGGAG CATCCCGAGT TGTGGGGTGG GCTGATCGAC CTGCCGGTGG
33241 GGGTTGATGA GGAGGTGTGC CGGCGGTTCG TGGGTGTTGT GGCGTCGGCT GGTTTTGAGG
33301 ATCAGGTGGC GGTGCGTGGT TCGGGTGTGT GGGTGCGTCG TCTGGTGCGT GCTGTGGTGG
33361 ATGGTGGTGG GGGTGGTTGG CGGCCGCGTG GGACGGTGTT GGTCACGGGT GGTCTTGGTG
33421 GTTTGGGTGC GCATACGGCC CGGTGGTTGG TGGGTGGTGG GGCGGATCAT GTGGTTCTTG
33481 TGAGCCGTCG TGGTGGCAGT GCGCCTGGTG CTGGGGATCT GGTGCGGGAG CTGGAGGGGT
33541 TGGGCGGGGC TCGGGTGTCG GTGCGGGCCT GTGATGTGGC TGATCGTGTG GCGTTGCGGG
33601 CGTTGTTGTC GGATCTGGGT GAGCCGGTGA CGGCGGTGTT CCATGCGGCT GGTGTTCCTC
33661 AGTCGACGCC TTTGGCGGAG ATCTCTGTCC AGGAGGCGGC TGATGTGATG GCGGCCAAGG
33721 TGGCGGGTGC GGTGAATCTG GGTGAGTTGG TGGATCCCTG TGGTCTGGAG GCGTTTGTGT
33781 TGTTCTCCTC CAATGCCGGT GTGTGGGGCA GTGGGGGGCA GGCGGTGTAT GCGGCGGCGA
33841 ATGCGTTTCT TGATGCGTTG GCGGTGCGTC GTCGGGGTGT TGGTCTGCCG GCCACGAGTG
33901 TGGCGTGGGG GATGTGGGCT GGTGAGGGGA TGGCGTCGGT GGGTGGTGCG GCGCGGGAGT
33961 TGTCCCGTCG GGGGGTGCGG GCGATGGATC CCGAGCGTGC TGTGGCGGTG ATGGCTGATG
34021 CGGTGGGTCG TGGTGAGGCG TTCGTCGCGG TCGCTGATGT GGACTGGGAA CGTTTCGTCA
34081 CCGGTTTCGC TTCTGCCCGT CCCCGTCCGT TGATCAGTGA CCTGCCGGAG GTGCGTGCTG
34141 TTGTGGAGGG CCAGGTCCAG GGCGGGGGCC AGGGGTTGGG CTTGGTCGGT GAGGAGGAGT
34201 CGTCGGGGTG GTTGAAGCGG TTGTCGGGGT TGTCTCGTGT GCGGCAGGAG GAGGAGTTGG
34261 TGGAGTTGGT CCGTGCTCAG GCTGCCGTTG TTCTCGGGCA TGGTTCCGCG CAGGACGTCC
34321 CGGCTGAGCG GGCGTTCAAG GAGTTGGGTT TTGATTCCCT CACTGCTGTC GAGCTACGCA
34381 ACGGGCTGGC CGCGGCCACC GGGATCCGGC TGCCGGCCAC CATGGCATTC GATCATCCCA
34441 ACGCCACCGC CATCGCACGC TTCCTGCAGT CTCAGCTCCT TCCTGACGCC GAGAGCGAGT
34501 CGGCCGTGCC GTCTTCACCG GAAGACGAGG TCCGCCAGGC ATTGGCGTCC CTTTCCCTGG
34561 ACCAGCTGAA AGGCGCTGGG CTTCTTGACC CACTGCTCGC TCTGACACGC CTCCGGGAGA
34621 TCAACAGCAC GGTGCAGAAC CCTGAGCCGA CCACCGAATC GATCGACGAG ATGGATGGCG
34681 AGACGTGCTG CGCCTGGCGC TCGGCGAAAT CGACGGCTGA GCCACTGACC ACTGGAGCTG
34741 ACATGCCTGA CCCCACCGCC AAATATGTGG AAGCGCTCCG TGCGTCGCTC AAGGAGAACG
34801 AACGCCTGCG CCAACAGAAT CACTCGCTTC TCGCCGCCTC CCGTGAAGCG ATCGCCATCA
34861 CGGCGATGAG CTGCCGTTTC GGCGGGGGCA TCGACTCGCC CGAAGATCTC TGGCGCTTCC
34921 TGGCCGAAGG CCGCGACGCG GTGGCGGGGC TTCCCGAGGA CCGCGGGTGG GATCTGGATG
34981 CCTTGTATCA CCCGGACCCG GAGAACCCCG GCACCACGTA CGTCCGGGAA GGCGCGTTCC
35041 GGTACGACGC AGCCCAGTTC GATGCGGGGT TCTTCGGGAT TTCGCCGCGT GAGGCGTTGG
35101 CGATGGACCC GCAGCAGCGG TTGCTGCTGG AGACATCCTG GGAGCTTTTC GAGCGTGCCG
35161 ATATCGATCC GTACACAGTC AGGGGAACGG CGACGGGGAT ATTCATCGGA GCCGGACATC
```

35221 AGGGCTATGG TCCCGACCCC AAGAGGGCTC CGGAGAGCGT GGCGGGTTAC CTGCTGACGG
35281 GAACGGCATC GGCCGTGCTG TCCGGGCGTA TTTCCTACAC GTTCGGTCTT GAGGGGCCTG
35341 CGGTCACGGT GGACACGGCG TGTTCGTCAT CGCTGGTGGC ACTGCACCTG GCGGTGCAGG
35401 CGCTGCGCCG GGGCGAGTGC TCACTCGCCA TAGCCGGCGG TGTGGCCGTC ATGTCGACCC
35461 CGGATGCCTT CGTGGAGTTC AGCCGCCAAC AGGGCATGGC AAGAGACGGC CGATGTAAGG
35521 CATTCGCCGC GGCAGCGGAC GGTATGGGAT GGGGCGAGGG AGTTTCGCTG CTCTTGCTGG
35581 AGCGGTTGTC GGATGCGCGG CGGTTGGGTC ATCGGGTGTT GGCGGTGGTG CGGGGGAGTG
35641 CGGTCAATCA GGATGGTGCG TCGAATGGCC TGGCGGCGCC GAATGGTCCG TCGCAGCAGC
35701 GGGTGATTCG TGCGGCGTTG GCTGACGCGG GTCTGGCTCC TGCCGATGTG GATGTGGTGG
35761 AGGCGCATGG TACGGGGACG CGGTTGGGTG ATCCGATCGA GGCTCAGGCG TTGCTGGCGA
35821 CGTATGGGCA GGGGCGTGCG GGTGGGCGTC CGGTGTGGCT GGGGTCGGTG AAGTCGAACA
35881 TCGGGCATAC GCAGGCGGCG GCTGGTGTGG CTGGTGTGAT GAAGATGGTG CTGGCGTTGG
35941 GGCGGGGTGT GGTGCCGAAG ACGTTGCATG TGGATGAGCC GTCACCGCAC GTGGACTGGT
36001 CGGCCGGTGC GGTGGAGTTG CTGACTGAAG AGCGGCCGTG GGAGCCGGAG GCTGAGCGTC
36061 TTCGTCGGGC AGGCATCTCC GCCTTCGGTG TCAGTGGCAC GAACGCGCAT GTGATCGTGG
36121 AGGAGGCGCC TGCGGAACCG GAGCCGGAGC CGGGAACTCG TGTGGTTGCT GCCGGTGATC
36181 TGGTGGTGCC GTGGGTGGTG TCCGGGCGGG ATGTGGGGGC GTTGCGTGAG CAGGCGGCAC
36241 GCTTGGCTGC GCACGTGTCG AGCACGGGTG CGGGTGTGGT TGATGTGGGC TGGTCGTTGG
36301 TGGCCACGAG GTCGGTGTTC GAGCACCGGG CGGTGATGGT CGGCACTGAT CTTGATTCCA
36361 TGGCGGGGTC GTTGGCCGGG TTTGCTGCGG GTGGTGTCGT CCCCGGGGTG GTGTCGGGTG
36421 TGGCGCCGGC TGAGGGTCGT CGTGTGGTGT TCGTCTTTCC TGGTCAGGGT TCGCAGTGGG
36481 TGGGGATGGC GGCTGGGTTG CTGGATGCGT GCCCGGTGTT CGCGGAGGCG GTGGCGGAGT
36541 GTGCCGCGGT GCTGGATCCG GTGACGGGTT GGTCGCTGGT CGAGGTGTTG CAGGGCAGGG
36601 ACGCGACTGT TCTTGGGCGG GTTGATGTGG TGCAGCCGGC GTTGTGGGCG GTGATGGTGT
36661 CACTGGCTCG GACCTGGCGG TATTACGGTG TGGAGCCTGC TGCGGTTGTG GGGCATTCGC
36721 AGGGTGAGAT TGCTGCGGCT TGTGTGGCTG GGGGGTTGAG TCTGGCCGAT GGTGCGCGGG
36781 TGGTGGTGTT GCGGAGCCGG GCGATCGCCC GGATCGCTGG TGGGGGCGGC ATGGTCTCCG
36841 TCAGTCTCCC GGCCGGCCGT GTCCGCACCA TGCTCGACAC CTACGGCGGC CGGGTTTCGG
36901 TCGCGGCGGT CAACGGTCCG TCCTCGACCG TGGTGTCCGG TGACGTCCAG GCCCTTGATG
36961 AGTTGTTGGC CGGTTGTGAG CGGGAGGGTG TCCGGGCTCG TCGTGTCCCG GTGGACTATG
37021 CCTCCCACTC CGCGCAGATG GACCAGTTAC GCGATGAGCT GCTGGAGGCG CTGGCGGACA
37081 TCACTCCGCA GGACTCCAGT GTTCCGTTCT TCTCGACGGT GACGGCGGAC TGGCTGGACA
37141 CGACCGCTCT GGATGCGGGG TACTGGTTCA CGAATCTGCG GGAGACGGTC CGGTTCCAGG
37201 AAGCCGTCGA AGGGCTTGTG GCTCAGGGGA TGGGCGCGTT CGTCGAGTGC AGCCCGCACC
37261 CCGTCCTCGT CCCCGGTATC GAGCAGACCC TCGACGCCCT CGACCAGAAT GCCGCCGTAC
37321 TCGGCTCGCT GCGGCGTGAC GAAGGCGGCC TGGACCGACT TCTCACATCC CTCGCGGAAG
37381 CCTTCGTCCA AGGCGTTCCC GTCGATTGGA CCCACGCCTT CGAGGGCGTG ACCCCTCGCA
37441 CCGTCGACCT GCCCACCTAC CCCTTCCAAC GGCAACGTTT CTGGTTGGAC GGTTCGCCGG
37501 CATCGTCTGC GAATGGCGTT GACGGTGAGG CGGACGCCAT GATCTGGGAC GCGGTCGAGC
37561 GTGAGGACTC GGTCGCTGTA GCCGAGGAGT TGGGGATCGA CGCCGAGGCT TTGCACACGG
37621 TGTTGCCGGC CTTGTCGTCG TGGCGGCGGC GTCGGGTGGA GCATCGACGG CTTCAGGACT
37681 GGCGTTACCG GGTGGAGTGG AAGCCTTTCC CGGCCGCGCT TGATGAGGTG CTCGGTGGTG
37741 GCTGGTTGTT CGTGGTGCCG CGGGGCTTGG CGGATGATGG TGTGGTTGCG CGGGTGGTGG
37801 CTGCCGTCAC GGCGCGGGGT GGCGAGGTCA GTGTCGTGGA GCTCGATCCG ACCCGTCCTG
37861 ACCGCCGGGC TTATGCGGAG GCTGTCGCGG GCCGTGGTGT GAGCGGGGTC GTGTCGTTCT
37921 TGTCCTGGGA TGATCGGCGG CACTCGGAGC ATCCTGTTGT TCCCGCCGGT CTTGCCGCGT
37981 CGCTGGTGTT GGCGCAGGCG TTGGTTGATC TTGGCCGGGT TGGTGAGGGG CCGCGGTTGT
38041 GGCTGGTGAC GCGGGATGCG GTGGTCGCTG GTCCTTCGGA TGCCGGTGCG GTGATTGATC
38101 CGGTACAGGC GCAGGTGTGG GGTTTCGGGC GTGTTCTGGG TCTGGAGCAT CCCGAGTTGT
38161 GGGGTGGGCT GATCGACCTG CCGGTGGAGG CGCCCGAACC TGGCTCGACG TGCGACCACA
38221 CGTATGCCGA CCTGCTCGCC ACGGTTGTGG CGTCGGCTGG TTTTGAGGAT CAGGTGGCGG
38281 TGCGTGGTTC GGGTGTGTGG GTGCGTCGTC TGGTGCGTGC TGTGGTGGAT GGTGGTGGGG
38341 GTGGTTGGCG GCCGCGTGGG ACGGTGTTGG TCACGGGTGG TCTTGGTGGT TTGGGTGCGC
38401 ATACGGCCCG GTGGTTGGTG GGTGGTGGGG CGGATCATGT GGTGCTTGTG AGCCGTCGTG
38461 GTGGCAGTGC GCCTGGTGCT GGGGATCTGG TGCGGGAGCT GGAGGGGTTG GGCGGGGCTC
38521 GGGTGTCGGT GCGGGCCTGT GATGTGGCTG ATCGTGTGGC GTTGCGGGCG TTGTTGTCGG
38581 ATCTGGGTGA GCCGGTGACG GCGGTGTTCC ATGCGGCTGG TGTTCCTCAG TCGACGCCTT
38641 TGGCGGAGAT CTCTGTCCAG GAGGCGGCTG ATGTGATGGC GGCCAAGGTG GCGGGTGCGG
38701 TGAATCTGGG TGAGTTGGTG GATCCCTGTG GTCTGGAGGC GTTTGTGTTG TTCTCCTCCA
38761 ATGCCGGTGT GTGGGGCAGT GGGGGGCAGG CGGTGTATGC GGCGGCGAAT GCGTTTCTTG
38821 ATGCGTTGGC GGTGCGTCGT CGGGGTGTTG GTCTGCCGGC GACGAGTGTG GCGTGGGGGA
38881 TGTGGGCTGG TGAGGGGATG GCGTCGGTGG GTGGTGCGGC GCGGGAGTTG TCCCGTCGGG
38941 GGGTGCGGGC GATGGATCCC GAGCGTGCTG TGGCGGTGAT GGCTGATGCG GTGGGGCGTG
39001 GTGAGGCGTT CGTCGCGGTC GCCGATGTGG ACTGGGAACG TTTCGTCACC GGTTTCGCCT
39061 CTGCCCGTCC CCGTCCGTTG ATCAGCGACC TCCCGGAGGT CCGTACCGCC CTGCGGAACC
39121 AGGAGCAGGA GCAACTCCAC GCCCCCGTCC CCGAGGACCG ATCGGCACAG CTTCTGCGGC
39181 GGCTGTCCAT GCTGTCTCCC GCCGGACGGG AAGCCGAACT GGTGAAGCTC GTCCGTACCG

```
39241 AGGCAGCCGC TGTTCTGGGG CACGGCTCCG CGCAGGACGT CCCGGCCGAG CGGGCGTTCA
39301 AGGAGCTGGG CTTCGACTCC CTCACCGCTG TTCAGCTACG CAACAGACTG GCCGCCGCCA
39361 CCGGCACCAG GCTCCCCGCC AGCGCCGTCT TCGACCACCC CCACGCTGCG GCTCTCGCCA
39421 GGTGGCTGCT CGCGGGGATG CGGCATGCCG ACGGTGGACA CGGTGGTGGG CACGCCGGTG
39481 GACCCGGGCC GGACGCCGAC GAAGGTCGGT CGGCCGGCGC TGGTCACAGC GGAATGCTGG
39541 CCGATCTGTA CCGGCGTTCC GCCGAGTTGG GCCGGAGCCG GGAGTTCATC GGGCTGCTGG
39601 CCGACACCGC GGCCTTCCGC CCGGTGTTCC ACGGGCCGGC GGACCTCGAC GCGCCGTTGG
39661 AGGCCGTTCC GCTGGCGGAC GGGGTGCGCA AACCGCAGTT GATCTGTTGC AGCGGGACCG
39721 CGCCGGTCGG CGGGCCGCAC GAGTTCGCGC GCCTGGCTTC GTTCTTCCGC GGCACTCGTG
39781 CGGTCTCGGC GCTTCCGCTG CCCGGCTACC TGCCCGGTGA GCAGTTGCCC GCGGACCTCG
39841 ACGCCGTGCT CGCCGCGCAG GCCGAGGCGG TCGAGAAGCA GACCGGGGGT GCGCCGTTCG
39901 TCCTGGTCGG CTACTCGGCG GGCGGACTGA TGGCCCACGC ACTGGCCTGC CACCTGGCCG
39961 GGCGCGGCAC ACCGCCGAGC GGTGAGGTGC TGGTGGACGT CTATCCGCCG GGCCGGCAGG
40021 AACCGGTGTT CGGCTGGCAG AAGGAGCTCA CCGAGGGCAT GTTCGCCCAG GACTTCGTGC
40081 CCATGGACGA TACGCGGCTG ACGGCCCTCG GCACGTACGA CCGTCTCATG GGCGAGTGGC
40141 GGCCGGCGCC CTCCGGACTG CCCACCCTCC TGATCCGGGC CACCGAACCC ATGGCGGAGT
40201 GGACCGGGGC CATCGACTGG CGGGCCTCCT GGGAGTACGA CCACACCGCC GTCGACATGC
40261 CGGGGAACCA CTTCACGATC ATGGCGCAGC ACGCGGAGGA CGCGGCCCGG CACATCGACG
40321 TCTGGCTGAA GGGGCTCACC CCCTGACACC TGCCCGCACC CTGTGACTCC TGCCCGTACC
40381 GGCGTCCCGG TCCTCCCGAC CCGCGTGCGC AACGGACGAG TCGCTCAGGA GGTCCCCATC
40441 GGCATGCCCC GCTTTCCTCC CCCTCTCCGA ACGCATCGAC GACCCGATCC CCCTCAGGGA
40501 CCGGTGAAGG AGCGTGTTGC ACTCATGCAG GACATGCAAG GCGTACAGCC CGAACCAGCC
40561 AGTGTCGAAC ACGCGGCGGA CGCAGCTCGA ACAGAGCGAA CGGCGCACGG AAGCCGCCCA
40621 GGAGATGGAG GACAGCGAAC TGGGGCGCCG CCTGCAGATG CTCCGCGGCA TGCAGTGGGT
40681 CTTCGGCGCC AACGGCGATC CGTACGCCCG GCTGCTGTGT GGCATGGAGG ATGACCCGTC
40741 ACCTTTCTAC GACGCGATAC GGACCCTGGG CGAGCTGCAC CGGAGCAGGA CCGGAGCCTG
40801 GGTCACCGCC GACCCCGGGC TCGGGGGCCG CATCCTCGCC GACCGGAAGG CTCGGTGCCC
40861 GGAAGGCTCG TGGCCGGTGC GGGCGAAGAC CGACGGGCTG GAGCAGTACG TGCTGCCCGG
40921 GCACCAGGCG TTCCTGCGGC TGGAGCGCGA GGAGGCCGAG CGACTGCGGG AGGTCGCGGC
40981 GCCGGTGCTG GGGGCCGCGG CGGTCGACGC GTGGCGCCCG CTGATCGACG AGGTCTGCGC
41041 GGGGCTCGCG AAGGGGCTGC CGGACACGTT CGACCTGGTC GAGGAGTACG CGGGGCTGGT
41101 GCCGGTCGAG GTGCTGGCGC GGATCTGGGG CGTCCCGGAG GAGGACCGCG CCCGGTTCGG
41161 GCGTGACTGC CGGGCGCTCG CTCCCGCGCT GGACAGCCTC CTGTGTCCCC AGCAGTTGGC
41221 GCTGAGCAAG GACATGGCGT CCGCCCTGGA GGACCTGCGT CTCCTCTTCG ACGGCCTCGA
41281 CGCGACGCCG CGCCTCGCCG GCCCCGCCGA CGGTGACGGA ACGGCCGTGG CCATGCTCAC
41341 CGTTCTGCTC TGCACGGAGC CGGTGACCAC GGCGATCGGG AACACCGTGC TCGGGCTCCT
41401 TCCCGGGCAG TGGCCCGTGC CCTGCACCGG CCGGGTGGCT GCCGGGCAGG TTGCCGGGCA
41461 GGCGCTGCAC CGGGCGGTGT CGTACCGTAT CGCGACGCGG TTCGCCCGGG AGGACCTGGA
41521 GTTGGCGGGC TGCGAGGTCA AGTCCGGTGA CGAGGTGGTG GTCCTGGCCG GAGCGATCGG
41581 CCGGAACGGA CCGTCCGCAG CCGCCCCGCC TGCCCCACCG GGCCCAGCGG CCCCGCCCGC
41641 CCCGTCGGTC TTCGGTGCCG CCGCCTTCGA GAACGCGCTG GCCGAACCCC TCGTCCGGGC
41701 TGTGACGGGA GCGGCCCTCC AGGCCCTCGC GGAGGGGCCC CCCCGGCTGA CGGCGGCGGG
41761 ACCCGTCGTA CGACGGCGGC GTTCCCCTGT CGTCGGCGGG CTGCACCGGG CTCCGGTGGC
41821 CGCCGCATGA GCATCGCGTC GAACGGCGCG CGCTCGGCCC CCCGCCGGCC CCTGCGCGTG
41881 ATGATGACCA CCTTCGCGGC CAACACGCAC TTCCAGCCGC TGGTTCCCCT GGCCTGGGCA
41941 CTGCGGACAG CCGGGCACGA GGTGCGCGTG GTGAGCCAGC CCTCGCTGAG CGACGTGGTG
42001 ACGCAGGCGG GGCTCACCTC GGTCCCGGTG GGCACCGAGG CTCCGGTCGA GCAGTTCGCG
42061 GCGACCTGGG GCGACGATGC CTACATCGGC GTCAACAGCA TCGACTTCAC CGGCAACGAC
42121 CCCGGCCTGT GGACGTGGCC GTACCTCCTG GGCATGGAGA CCATGCTGGT GCCGGCCTTC
42181 TACGAGTTGC TGAACAACGA GTCCTTCGTG GACGGCGTAG TCGAGTTCGC CCGTGACTGG
42241 CGGCCCGACC TGGTGATCTG GGAGCCGCTG ACGTTCGCCG GCGCGGTGGC GGCGCGCGTC
42301 ACCGGCGCGG CCCACGCCCG GCTGCCGTGG GGGCAGGAGA TCACCCTGCG CGGGCGGCAG
42361 GCGTTCCTCG CCGAGCGTGC CCTGCAACCG TTCGAGCACC GGGAGGATCC CACGGCCGAG
42421 TGGCTGGGCC GCATGCTCGA CCGGTACGGC TGCTCGTTCG ACGAGGAGAT GGTCACCGGG
42481 CAGTGGACCA TCGACACGCT GCCGCGCAGC ATGCGGCTGG AGCTGTCCGA GGAGCTGCGC
42541 ACCCTGGACA TGCGGTACGT GCCGTACAAC GGACCGGCGG TCGTACCCCC CTGGGTGTGG
42601 GAACCGTGCG AGCGGCCCCG GGTCTGTCTG ACGATCGGCA CCTCCCAGCG TGACTCCGGC
42661 CGGGACCATG TCCCCCTCGA CCACCTGCTC GACTCCCTCG CCGACGTGGA CGCGGAGATC
42721 GTGGCCACGC TCGACACCAC CCAGCAGGAG CGCCTGCGGG GCGCGGCCCC CGGCAACGTC
42781 CGGCTGGTGG ACTTCGTCCC GCTGCACGCG CTGATGCCGA CCTGCTCGGC GATCGTGCAC
42841 CACGGTGGTC CGGGCACGTG GTCGACGGCG GCGCTCCACG GCGTCCCGCA GATCATCCTG
42901 GACACCTCGT GGGACACACC GGTGCGGGCG CAGCGCATGC AGCAACTCGG GGCGGGCCTG
42961 TCGATGCCGG TGGGGGAACT GGGCGTCGAG GCGCTGCGGG ACCGGGTCCT GCGGCTGCTG
43021 GGGGAGCCGG AGTTCCGCGC GGGCGCCGAG CGGATCCGGG CCGAGATGCT CGCGATGCCC
43081 GCCCCCGGTG ACGTCGTACC GGACCTGGAA CGACTCACCG CGGAGCATGC CACCGGCGCG
43141 ATGGCGGGAA GGCGGTGAGA CGATGCGCGT ACTGCTGACC TGCTTCGCCA ACGACACCCA
43201 CTTCCACGGG CTGGTGCCGC TGGCGTGGGC GCTGCGGGCC GCCGGGCACG AAGTCCGCGT
```

```
43261 GGCCAGTCAG CCCGCCCTGT CCGACACGAT CACCCAAGCG GGACTGACCG CGGTGCCCGT
43321 GGGCCGGGAC ACCGCCTTCC TGGAGCTGAT GGGGGAGATC GGCGCGGACG TCCAGAAGTA
43381 CTCCACCGGC ATCGACCTGG GCGTCCGCGC GGAGCTGACG AGCTGGGAGT ACCTGCTCGG
43441 CATGCACACG ACCCTGGTGC CCACGTTCTA CTCGCTGGTC AACGACGAGC CGTTCGTCGA
43501 CGGGCTCGTC GCGCTGACCC GGGCCTGGCG GCCCGACCTC ATCCTGTGGG AGCACTTCAG
43561 CTTCGCCGGG GCGTTGGCGG CGCGGGCCAC CGGCACGCCC CACGCCCGCG TGCTGTGGGG
43621 GTCGGACCTC ATCGTCCGGT TCCGCCGGGA CTTCCTCGCG GAGCGGGCGA ACCGGCCCGC
43681 CGAGCACCGC GAGGACCCCA TGGCGGAGTG GCTGGGCTGG GCGGCCGAAC GGCTGGGCTC
43741 CACCTTCGAC GAGGAGCTGG TGACCGGGCA GTGGACGATC GACCCGCTGC CGCGGAGCAT
43801 GCGGCTGCCC ACCGGGACGA CGACGGTGCC GATGCGGTAC GTGCCGTACA ACGGGCGGGC
43861 CGTGGTCCCC GCATGGGTCC GGCAGCGTGC GCGGCGGCCC CGGATCTGCC TGACGCTCGG
43921 TGTGTCGGCC CGGCAGACCC TGGGCGACGG CGTGTCGCTG GCGGAGGTGC TGGCCGCGCT
43981 GGGCGACGTG GACGCGGAGA TCGTGGCCAC GCTGGACGCC TCCCAGCGCA AGCTCCTGGG
44041 GCCGGTGCCG GACAACGTCC GGCTGGTGGA CTTCGTGCCC CTGCACGCCC TGATGCCGAC
44101 CTGTTCGGCG ATCGTGCACC ACGGCGGCGC CGGTACCTGG CTGACGGCCG CCGTCCACGG
44161 CGTCCCGCAG ATCGTCCTCG GTGACCTCTG GGACAACCTG CTGCGCGCCC GGCAGACACA
44221 GGCCGCGGGC GCGGGCCTGT TCATCCATCC GTCCGAGGTC ACCGCGGCCG GGCTCGGTGA
44281 GGGCGTGCGC CGGGTGCTGA CGGACCCTTC CATCCGGGCC GCCGCACAGC GCGTCCGGGA
44341 CGAGATGAAT GCAGAGCCGA CGCCGGGCGA GGTCGTCACG GTGCTGGAGC GGCTCGCCGC
44401 GAGCGGCGGA CGCGGACGAG GAGGCGGGAA CCATGCGGGC TGACACGGAG CCGACCACCG
44461 GGTACGAGGA CGAGTTCGCC GAGATCTACG ACGCCGTGTA CCGGGGCCGG GGCAAGGACT
44521 ACGCCGGCGA GGCGAAGGAC GTGGCGGACC TCGTGCGCGA CCGGGTGCCG GACGCGTCCT
44581 CCCTCCTGGA CGTGGCCTGC GGCACGGGCG CGCACCTGCG GCACTTCGCC ACGCTCTTCG
44641 ACGACGCCCG CGGTCTCGAA CTGTCCGCGA GCATGCTGGA CATCGCCCGC TCCCGCATGC
44701 CGGGCGTGCC GCTGCACCAA GGGGACATGC GATCCTTCGA CCTGGGGCCA CGCGTCTCCG
44761 CGGTCACCTG CATGTTCAGC TCCGTCGGCC ACCTGGCCAC CACCGCCGAA CTCGACGCGA
44821 CGCTGCGGTG CTTCGCCCGG CACACCCGGC CCGGCGGCGT GGCCGTCATC GAACCGTGGT
44881 GGTTCCCGGA GACCTTCACC GACGGCTACG TGGCGGGTGA CATCGTACGC GTCGACGGCC
44941 GGACCATCTC CCGGGTGTCC CACTCGGTAC GGGACGGCGG CGCCACCCGC ATGGAGATCC
45001 ACTACGTGAT CGCCGACGCC GAGCACGGTC CCCGGCACCT GGTCGAGCAC CACCGCATCA
45061 CGCTGTTCCC GCGGCATGCG TACACGGCCG CGTACGAGAA GGCGGGCTAC ACCGTCGAGT
45121 ACCTCGACGG CGGGCCCTCG GGCCGGGGGC TGTTCGTCGG CACCCGGACG TGAACCCGCC
45181 CGCGCACCGC CCGATCACCC TGCTCAACGC CGTTCACACG GATCACCGGA CCACGCGAAG
45241 GACCTTTCAC ATGTCGTACG ACGACCACGC GGTGCTGGAA GCGATACTGC GGTGCGCCGG
45301 AGGTGACGAG CGCTTCCTGC TGAACACCGT CGAGGAATGG GGAGCCGCCG AGATCACCGC
45361 GGCGCTCGTG GACGAGTTGC TGTTCCGCTG CGAGATCCCG CAGGTGGGCG GTGAGGCGTT
45421 CATCGGCCTG GACGTCCTGC ACGGCGCCGA CCGGATCAGC CATGTGCTGC AGGTGACGGA
45481 CGGCAAGCCG GTCACGTCGG CGGAACCGGC CGGCCAGGAA CTGGGCGGCC GTACCTGGAG
45541 TTCACGCTCA GCGACCCTCC TGCGGGAGCT GTTCGGCCCG CCGTCCGGCC GCACCGCGGG
45601 GGGCTTCGGC GTCTCCTTCC TGCCCGACCT GCGCGGCCCG CGGACCATGG AGGGCGCGGC
45661 CCTGGCCGCC CGCGCCACCA ACGTGGTGCT GCACGCGACG ACCAACGAGA CGCCCCCACT
45721 GGACCGGCTG GCCCTGCGCT ACGAGTCCGA CAAGTGGGGC GGCGTCCACT GGTTCACCGG
45781 CCACTACGAC CGGCACCTGC GGGCCGTGCG CGACCAGGCG GTGCGGATCC TGGAGATCGG
45841 CATCGGCGGC TACGACGACC TGCTGCCGAG CGGCGCCTCA CTGAAGATGT GGAAGCGCTA
45901 CTTCCCGCGC GGCCTGGTCT TCGGCGTGGA CATCTTCGAC AGTCGGCGTG CGACCAGCCG
45961 CGTGTCAAGA CGCTCCGCGG CCCGGCAGGA CGACCCGGAG TTCATGCGCC GCGTCGCCGA
46021 GGAGCACGGG CCGTTCGACG TCATCATCGA CGACGGCAGC CACATCAACG CACACATGCG
46081 GACGTCGTTC TCGGTGATGT TCCCCCACCT GCGCAACGGC GGCTTCTACG TCATCGAGGA
46141 CACCTTCACC TCCTACTGGC CCGGGTACGG AGGGCCATCC GGAGCCCGGT GCCCGTCCGG
46201 AACAACCGCG CTGGAGATGG TCAAGGGACT GATCGACTCG GTGCACTACG AGGAGCGGCC
46261 GGACGGCGCG GCCACGGCCG ACTACATCGC CAGGAACCTC GTCGGGCTGC ACGCCTACCA
46321 AACGACCTCG TCTTCCTCGA GAAGGGCGAT CAACAAGGAG GGCGGCATCC CCCACACCGT
46381 GCCCCGGGAG CCGTTCTGGA ACGACAACTA GCCACGGCCG CAACCAGAGC CGGAAACCGC
46441 ACCACTGTCC GCGCCACCTC GGAACCACCT CCAGCAAAGG ACACACCGCT GTGACCGATA
46501 CGCACACCGG ACCGACACCG GCCGACGCGG TACCCGCCTA CCCGTTCAGC CTGCCGCACG
46561 CCCTGGACCT CGACCCGCAC TACGCCGAAC TCCGCCGCGA CGAACCCGTC TCCAGGGTGC
46621 GCCTGCCCTA CGGCGAGGGC ACGGCCTGGC TGGTCACCCG CATGTCCGAC GCCCGTATCG
46681 TTCTGGGCGA CTCCCGCTTC AGCACCGCGG CCGCCACCGA TCCCGCCACC CCCCGGATGT
46741 TCCCCACCCC GCCCGAGCCG GACGGCGTCC TCGCCCAGGA CCCGCCGGAC CACACCCGGC
46801 TGCGGCGGCT GGTGGGCAAG GCCTTCACGG CACGCCGGGT GGAGGAGATG CGGCCCCGTG
46861 TCCGCTCCCT CGTCGACTCC CTGCTCGACG ACATGGTGGC GCACGGTTCA CCCGCCGACC
46921 TGGTCGAGTT CCTCGCCGTT CCCTTCCCCG TCGCGGTCAT CTGCGAACTG CTCGGCGTGC
46981 CCTTGGAGGA CCGCGACCTG TTCCGGACCT TCTCCGACGC CATGCTCTCC TCGACCCGGC
47041 TCACCGCCGC GGAGATACAG CGGGTCCAGC AGGACTTCAT GGTCTACATG GACGGCCTGG
47101 TCGCCCAGCG CCGCGACGCC CCCACCGAGG ACCTGCTCGG CGCCCTCGCC CTCGCCACCG
47161 ACAACGACGA CCACCTGACC AAGGGCGAGA TCGTCAACAT GGGGGTGAGC CTGCTCATCG
47221 CGGGCCACGA GACGTCGGTC AACCAGATCA CCAACCTCGT CCACCTCCTG CTGACCGAGC
```

```
47281 GCAAGCGCTA CGAGTCGCTG GTCGCCGACC CGGCCCTCGT GCCCGCGGCG GTGGAGGAGA
47341 TGCTGCGGTA CACACCGCTG GTGTCCGCCG GCAGCTTCGT CCGCGTGGCC ACCGAGGACG
47401 TGGAGCTGAG CACCGTGACC GTGCGGGCCG GGGAGCCCTG CGTCGTCCAC TTCGCGTCGG
47461 CCAACCGGGA CGAGGAGGTC TTCGACCACG CCGACGAGCT GGACTTCCAC CGTGAGCGCA
47521 ACCCGCACAT AGCGTTCGGG CACGGAGCGC ACCACTGCAT CGGCGCCCAA CTGGGCCGAC
47581 TGGAACTCCA GGAGGCCCTG TCCGCCCTCG TCCGGCGCTT CCCCACCCTC GATCTGGCCG
47641 AGCCGGTCGC GGGACTGAAG TGGAAGCAGG GCATGCTGAT CCGCGGACTG GAACGCCAGA
47701 TCGTCTCCTG GTGACGGCCG GCCGCCCGGC CGCCCGCCGG GCACCGGCGC CACCAGGGCA
47761 CCGGCCGGGA CCGCAGACCC GGCCGGTGCC CCTCGCCCGA GGCCGCCTCA CTCCACGAAG
47821 CGGCCACCCT CCATGTGCAT GCGGCGACCG GTGAACCGCT GCGCGAACAT GCGGTCGTGG
47881 GAGACCACGA CCAGTGCGCC CCGGTAGTGC GCCAGCGCCT CCTCCAGGTC CTCCACGAGC
47941 GCGGGCGACA GGTGGTTCGT CGGCTCGTCG AGCAGCAGCA GGTCCGCCGG GTCGCGCAGC
48001 AGACGGGCCA GGGCCAGCCG CCTCAACTGC CCGGTGGACA GGTCTCCCAC CGCGGTGCCC
48061 AGCGCCGAGG GCCGGAAGAG CCCGAATCCC AGGAGCGCGC CCCGGTGTTC CTCCGCGATG
48121 CCGGGCAGCC CCGCCGCGAA GGCCGCCAGC AGGCTCTGCT GCCGGTCGGT GATCTCCGTC
48181 TCCTGCGGCA GCCAGCCGAT GCGCTCCGGG CGCTCGCACT CGCCCTGATC GGGCGCCAGG
48241 TCACCGGCCA GCACGCGCAG CAGGGTGCTC TTGCCCGCGC CGTTGTGCCC CGTGATCAGG
48301 ATGCGCTCAC CGGGGTCGAC GGTGAAGGAC GGGACGTCGA GCCGCGTGCC GACGGTGACC
48361 TTGTACAGCT CGGCGAGTGC CCCGCCGCGC CCGACCGTGC CGCCACCCTC CACCCGGGCC
48421 CGGAAACGCA TGGGTTGAGG GGGCCGCGGC ACCGGGTTCT CCTCCAGCCG GCGGACCCGC
48481 TCCTTGGCGT TGCGGACCCG CGCGGAGATC TGCTTCTCCA CGTTGCGCTG GTGGCGCTGG
48541 TTCGACCGCT CGGTGTTGCG CCGCGGGCCG GTGGCCAGGT GGTCGGCGGC GCTGCGGGCC
48601 AGTTCCCGCT GGCGTGCCAG GTCCTCCAGC CAGTCCTGGT AAGCCTGCTC CCAGCGGCGC
48661 CGCGCGGCCG CCTTGGCTTG CAGGTATCCC GCGTAACCGC CGCCGTGCCG GTTGACGGTG
48721 CGCCGCTCGC CGTCCACCTC CCACAGGGCG GTGGCCACGC GCTCCAGGAA GACCCGGTCG
48781 TGCGAGACGA CCAGCACGCT GCCGCGGTGG GCCCGCAGGC GCTCCTCCAG CCACTCCAGC
48841 GCCCCGACGT CGAGGTGGTT GGTGGGTTCG TCGAGCAGCA TCAGCTGCGG GGACGCGGCC
48901 AGCAGGCAGG CCAGGTTGAG ACGCGCCTGC TCACCTCCGG AGAGGCTGCC GAGCCGCCGG
48961 TCGCCCGTGA TGCCCGCCAG ACCGAGGCCG TGCATCGCCG CGTCGACACG GGCGTCCGCC
49021 GCGTAGCCGT CGCGGGCCTC GAACGCCTCC AGCAGGTCGC CGTAGGCGCC GAGCAGGCCC
49081 TCCAGCTCCT CGGGCTCCGC CCCGGCCAGC GCCTGCTCCG CCTCACGCAA CCCCCGCTCC
49141 AGGGAGCGCA GTTCGGCGAG GGCGTGGTCG ATGGCGTCCT GAACGGTGTC CTCCGGGGGC
49201 AGGTCCGGTG TCTGGGGGAG GTAGCCGCAG CCGCCGGGAG CCCGGACGAG GACCTGGCCA
49261 CCGTCCGGGC GGTCCACGCC GGCGAGCATG CGGAGCAGGG TCGACTTGCC CGATCCGTTC
49321 TCACCGATGA TGCCGACGCG CTCGCCGAGT GCCACCGACT GGTTGACGCC GTCCAACAGC
49381 GGCCGTCCGC CGGGTGCCCG GACGACGTCG TCGAGGACGA CCTGGAAGGA ACCGGTCTCA
49441 GGTTGCGTGG GAAGGAGCTT TTCCGGCGTG CCGGTGAGCG CCGCGGCGCC GGTATCGGAA
49501 CGGTGTGCGT TCTGCATGGG TGATCCGCCA TTCGGAGAAA AAGAGGCAGT GTGGCCAAAA
49561 GGGAGCGGCC CACGGCAGAC GGCGGAAGAA GAGAACGCCT CGGCGAACGC GGCGCACCCG
49621 ACGGTGCGCA GCGCGAAAAA AGGGAGGCGA AGAAGCGAGC CGGAGGCGTC GCGATCAGCG
49681 GCGGGAGAAG CCGCGTCACC GTCCTGCCGG GAACCCTCGA CGGCGCCGGA GCGGCAACCG
49741 CGTACGCGGT GCTCCTCGGC GCCCGGACTC CCGTGGCGGT ATCAGAGGAA GTAGTAACTG
49801 ACCACGTCGG CACGATAGCA GAGCAGACGG AGCCGGCAGG GGGTCGCGAG GTGCGATGGC
49861 TGAATGTGTG CCACGCTTCG GATTTTTTGC TCGCGGGACG ACGAGGCCGT GTGCGAACGT
49921 GTCCCGGGCA GTCGTTCGTC AGCGGGAGGT TCATATGCAG GACAACCAGG GTGGATCCGG
49981 AGCCGAGTCC GAGACCGGGA CCGAGAGCGA CGTCAAGCGG AAGTTCCGGG AGGCACTGGA
50041 GCGCAAGAAG CTCCTCAGCC GGGAACGCCG GGCGCACGAG GACGCTCGTT CCAAGGTGAA
50101 CGGAACGTCC CGCAATGGCG CCAGGAAGGC GAATTTCCGC CGCAAGGCCG GGTGACACCG
50161 ACCGCTGCGC ACACCCGTGC CCCACAGCTC GACTCCGCTG CGACAGGGGC CTGCCCGCGC
50221 CGGGGAACCG GCCCGGGCAG GTGTAGGGTG GCGGGCATGT ATCCAGGTGT CGGTTCCCTG
50281 AAGCTCCGCC GCCGCGCCTG ACGGTGCGGC CCTGAACTCT CGTTTCGCGT GCCCACCGTC
50341 GCGGTGTCAG TGCCGGGCGG CTGTTTCGTG CTGCCCGGTT CCGGAGCGAA CCTGTGGAGC
50401 ACACCGTGGG CGCATTCCCC GCAAGGCCGG CCTGAGGCCG CGACCGATAC ACGAGTTCAC
50461 CGATGCGAGC GAGGGCCGCC GCCGCGCCGG TGGCGACGAC CACCCCTTCC GCACCGGCCC
50521 CGACGCCCTC TCGCCGGCGC CGCTCCCGGC CCCGGCCGGC GGCGCCACCC GGGTACGCCG
50581 CTCCCGCGGC CCCGGCGGCG CGTGCCGCGC ACAGGCCGTA CCGGCCGGCC GTTCGGCCGG
50641 TGGACCTCTG CGCCCTGCCG TCCCCGCGGC AACGTCGCCG GACACGGACA CCGCCCCTCG
50701 GCCGCCGGCC GCCGTCACCA CCCCGGGGCG CCGGCGTCTC GCCGCTCTCG CGCCGGCCCC
50761 GTCCACGACC GCTCCCGTGC CTGCCGGAAG GGCCGACTCA TGACCGAGCG ACACCTCCCC
50821 GCCGTCCTCG CGCCCCTCGG CCGGCCGGGC TACCGCCGCC TCTTCGCCGC CATGGTCCTC
50881 GCCCTCTTCG GGTACGGCGG GTGGACCATC TACCTCGCGC TCCAGGCGCT GGAGCTC
```

The above DNA sequence encodes the following 8,8a-deoxyoleandolide synthase proteins:

8,8a-deoxyoleandolide synthase 1:

```
   1 MHVPGEENGH SIAIVGIACR LPGSATPQEF WRLLADSADA LDEPPAGRFP TGSLSSPPAP
  61 RGGFLDSIDT FDADFFNISP REAGVLDPQQ RLALELGWEA LEDAGIVPRH LRGTRTSVFM
 121 GAMWDDYAHL AHARGEAALT RHSLTGTHRG MIANRLSYAL GLQGPSLTVD TGQSSSLAAV
 181 HMACESLARG ESDLALVGGV NLVLDPAGTT GVERFGALSP DGRCYTFDSR ANGYARGEGG
 241 VVVVLKPTHR ALADGDTVYC EILGSALNND GATEGLTVPS ARAQADVLRQ AWERARVAPT
 301 DVQYVELHGT GTPAGDPVEA EGLGTALGTA RPAEAPLLVG SVKTNIGHLE GAAGIAGLLK
 361 TVLSIKNRHL PASLNFTSPN PRIDLDALRL RVHTAYGPWP SPDRPLVAGV SSFGMGGTNC
 421 HVVLSELRNA GGDGAGKGPY TGTEDRLGAT EAEKRPDPAT GNGPDPAQDT HRYPPLILSA
 481 RSDAALRAQA ERLRHHLEHS PGQRLRDTAY SLATRRQVFE RHAVVTGHDR EDLLNGLRDL
 541 ENGLPAPQVL LGRTPTPEPG GLAFLFSGQG SQQPGMGKRL HQVFPGFRDA LDEVCAELDT
 601 HLGRLLGPEA GPPLRDVMFA ERGTAHSALL SETHYTQAAL FALETALFRL LVQWGLKPDH
 661 LAGHSVGEIA AAHAAGILDL SDAAELVATR GALMRSLPGG GVMLSVQAPE SEVAPLLLGR
 721 EAHVGLAAVN GPDAVVVSGE RGHVAAIEQI LRDRGRKSRY LRVSHAFHSP LMEPVLEEFA
 781 EAVAGLTFRA PTTPLVSNLT GAPVDDRTMA TPAYWVRHVR EAVRFGDGIR ALGKLGTGSF
 841 LEVGPDGVLT AMARACVTAA PEPGHRGEQG ADADAHTALL LPALRRGRDE ARSLTEAVAR
 901 LHLHGVPMDW TSVLGGDVSR VPLPTYAFQR ESHWLPSGEA HPRPADDTES GTGRTEASPP
 961 RPHDVLHLVR SHAAAVLGHS RAERIDPDRA FRDLGFDSLT ALELRDRLDT ALGLRLPSSV
1021 LFDHPSPGAL ARFLQGDDTR RPEPGKTNGT RATEPGPDPD DEPIAIVGMA CRFPGGVTSP
1081 EDLWRLLAAG EDAVSGFPTD RGWNVTDSAT RRGGFLYDAG EFDAAFFGIS PREALVMDPQ
1141 QRLLLETSWE ALERAGVSPG SLRGSDTAVY IGATAQDYGP RLHESDDDSG GYVLTGNTAS
1201 VASGRIAYSL GLEGPAVTVD TACSSSLVAL HLAVQALRRG ECSLALAGGA TVMPSPGMFV
1261 EFSRQGGLSE DGRCKAFAAT ADGTGWAEGV GVLLVERLSD ARRLGHRVLA VVRGSAVNQD
1321 GASNGLTAPN GPSQQRVIRA ALADAGLVPA DVDVVEAHGT GTRLGDPIEA QALLATYGQG
1381 RAGGRPVVLG SVKSNIGHTQ AAAGVAGVMK MVLALGRGVV PKTLHVDEPS AHVDWSAGEV
1441 ELAVEAVPWS RGGRVRRAGV SSFGISGTNA HVIVEEAPAE PEPEPERGPG SVVGVVPWVV
1501 SGRDAGALRE QAARLAAHVS GVSAVDVGWS LVATRSVFEH RAVMVGSELD AMAESLAGFA
1561 AGGVVPGVVS GVAPAEGRRV VFVFPGQGSQ WVGMAAGLLD ACPVFAEAVA ECAAVLDPLT
1621 GWSLVEVLRG GGEAVLGRVD VVQPALWAVM VSLARTWRYY GVEPAAVVGH SQGEIAAACV
1681 AGGLSLADGA RVVVLRSRAI ARIAGGGGMV SVSLPAGRVR TMLEEFDGRV SVAAVNGPSS
1741 TVVSGDVQAL DELLAGCERE GVRARRVPVD YASHSAQMDQ LRDDLLEALA TIVPTSANVP
1801 FFSTVTADWL DTTALDAGYW FTNLRETVRF QEAVEGLVAQ GMGAFVECSP HPVLVPGITE
1861 TLDTFDADAV ALSSLRRDEG GLDRFLTSLA EAFVQGVPVD WSRAFEGASP RTVDLPTYPF
1921 QRQRYWLLDK AAQRERERLE DWRYHVEWRP VTTRPSARLS GVWAVAIPAR LARDSLLVGA
1981 IDALERGGAR AVPVVVDERD HDRQALVEAL RNGLGDDDLA GVLSLLALDE APHGDHPDVP
2041 VGMAASLALV QAMADAAAEV PVWFATRGAV AALPGESPER PRQALLWGLG RVVALEQPQI
2101 WGGLVDLPQH LDEDAGRRLV DVVGGLADED QLAVRASSVL ARRLVRTPGH RMSSQAGGRE
2161 WSPSGTVLVT GGTGALGAHV ARWLAGKGAE HLVLISRRGA DAAGAAALRD SLTDMGVRVT
2221 LAACDAADRH ALETLLDSLR TDPAQLTAVI HAAGALDDGM TTVLTPEQMN NALRAKVTAT
2281 VNLHELTRDL DLSAFVLFSS ISATLGIPGQ ANYAPGNSFL DAFAEWRRAQ GLVATSIAWG
2341 PWSGGTGMAH EGSVGERLQR HGVLAMEPAA AIAALDHTLA SDETAVAVAD IDWSRFFLAY
2401 TALRARPLIG EIPEARRMLE SGSGPGDLEP DRAEPELAVR LAGLTAVEQE RLLVQLVREQ
2461 AAVVLGHSGA EAVAPDRAFK DLGFDSLTSV ELRNRLNTAT GLRLPVTAVF DYARPAALAG
2521 HLRSRLIDDD GDHGALPGVE KHAIDEPIAI VGMACRFPGG IASPEDLWDV LTAGEDVVSG
2581 LPQNRGWDLG RLYDPDPDRA GTSYMREGAF LHEAGEFDAA FFGISPREAL AMDPQQRLLL
2641 ETSWEALERA GITPSKLAGS PTGVFFGMSN QDYAAQAGDV PSELEGYLLT GSISSVASGR
2701 VAYTFGLEGP AVTVDTACSS SLVALHLAVQ GLRRGECSLA LVGGVTVMSS PVTLTTFSRQ
2761 RGLSVDGRCK AFAASADGFG AAEGVGVLLV ERLSDARRLG HRVLAVVRGS AVNQDGASNG
2821 LAAPNGPSQQ RVIRAALADA GLAPADVDVV EAHGTGTRLG DPIEAQALLA TYGQGRTSGR
2881 PVWLGSVKSN IGHTQAAAGV AGVMKMVLAL GRGVVPKTLH VDEPSPHVDW SAGEVELAVE
2941 AVPWSRGGRV RRAGVSSFGI SGTNAHVIVE EAPAEPSVEE GPGSVVGVVP WVVSGRDAGA
3001 LRAQAARLAA HVSSTGAGVV DVGWSLVATR SVFEHRAVMV GTDLDSMAGS LAGFAAGGVV
3061 PGVVSGVAPA EGRRVVFVFP GQGSQWVGMA AGLLDACPVF AEAVAECAAV LDRLTGWSLV
3121 EVLRGGEAVL GRVDVVQPAL WAVMVSLART WRYYGVEPAA VVGHSQGEIA AACVAGGLSL
3181 ADGARVVVLR SRAIARIAGG GGMVSVGLSA ERVRTMLDTY GGRVSVAAVN GPSSTVVSGD
3241 AQALDELLAG CEREGVRARR VPVDYASHSA QMDQLRDELL EALADVTPQD SSVPFFSTVT
3301 ADWLDTTALD AGYWFTNLRE TVRFQEAVEG LVAQGMGAFV ECSPHPVLVP GITETLDTFD
3361 ADAVALSSLR RDEGGLDRFL TSLAEAFVQG VPVDWTHAFE GGRPRFVDLP TYAFQRQRYW
3421 LHEEPLQEPV DEAWDAEFWS VVERGDATAV SDLLSTDAEA LHTVLPALSS WRRRRVEHRR
3481 LQDWRYRVEW KPFPAALDEV LGGGWLFVVP RGLADDGVVA RVVAAVTARG GEVSVVELDP
3541 TRPDRRAYAE AVAGRGVSGV VSFLSWDDRR HSEHSVVPAG LAASLVLAQA LVDLGRVGEG
3601 PRLWLVTRGA VVAGPSDAGV VIDPVQAQVW GFGRVLGLEH PELWGGLVDL PVGVDEEVCR
3661 RFVGVVASAG FEDQVAVRGS GVWVRRLVRA VVDGGGGGWR PRGTVLVTGG LGGLGAHTAR
3721 WLVGGGADHV VLVSRRGGSA PGAGDLVREL EGLGGARVSV RACDVADRVA LRALLSDLGE
3781 PVTAVFHAAG VPQSTPLAEI SVQEAADVMA AKVAGAVNLG ELVDPCGLEA FVLFSSNAGV
3841 WGSGGQAVYA AANAFLDALA VRRRGVGLPA TSVAWGMWAG EGMASVGGAA RELSRRGVRA
3901 MDPERAVAVM ADAVGRGEAF VAVADVDWER FVTGFASARP RPLISDLPEV RAVVEGQVQG
```

```
3961 RGQGLGLVGE EESSGWLKRL SGLSRVRQEE ELVELVRAQA AVVLGHGSAQ DVPAERAFKE
4021 LGFDSLTAVE LRNGLAAATG IRLPATMAFD HPTATAIARF LQSELVGSDD PLTLMRSAID
4081 QLETGLALLE SDEEARSEIT KRLNILLPRF GSGGSSRGRE AGQDAGEHQD VEDATIDELF
4141 EVLDNELGNS
```

8,8a-deoxyoleandolide synthase 2:

```
   1 VTNDEKIVEY LKRATVDLRK ARHRIWELED EPIAITSMAC HFPGGIESPE QLWELLSAGG
  61 EVLSEFPDDR GWDLDEIYHP DPEHSGTSYV RHGGFLDHAT QFDTDFFGIS PREALAMDPQ
 121 QRLLLETSWQ LFERAGVDPH TLKGSRTGVF VGAAHMGYAD RVDTPPAEAE GYLLTGNASA
 181 VVSGRISYTF GLEGPAVTVD TACSSSLVAL HLAVQALRRG ECSLAVVGGV AVMSDPKVFV
 241 EFSRQRGLAR DGRSKAFAAS ADGFGFAEGV SLLLLERLSD ARRLGHRVLA VVRGSAVNQD
 301 GASNGLAAPN GPSQQRVIRA ALADAGLAPA DVDVVEAHGT GTRLGDPIEA QALLATYGQG
 361 RTSGRPVWLG SVKSNIGHTQ AAAGVAGVMK MVLALERGVV PKTLHVDEPS PHVDWSTGAV
 421 ELLTEERPWE PEAERLRRAG ISAFGVSGTN AHVIVEEAPA EPEPEPEPGT RVVAAGDLVV
 481 PWVVSGRDAG ALRAQAARLA AHVSSTGAGV VDVGWSLVAT RSVFEHRAVM VGTDLDSMAG
 541 SLAGFAAGGV VPGVVSGVAP AEGRRVVFVF PGQGSQWVGM AAGLLDACPV FAEAVAECAA
 601 VLDPLTGWSL VEVLRGGEAV LGRVDVVQPA LWAVMVSLAR TWRYYGVEPA AVVGHSQGEI
 661 AAACVAGGLS LADGARVVVL RSRAIARIAG GGGMVSVSLP AGRVRTMLDT YGGRLSVAAV
 721 NGPSSTVVSG DAQALDELLA GCEREGVRAR RVPVDYASHS AQMDQLRDEL LEALADITPQ
 781 HSSVPFFSTV TADWLDTTAL DAGYWFTNLR ETVRFQEAVE GLVAQGMGAF VECSPHPVLV
 841 PGIEQTLDTV EADAVALGSL RRDEGGLGRF LTSLAEAFVQ GVPVDWSRTF EGASPRTVDL
 901 PTYPFQRQRF WLEGSPALSS NGVEGEADVA FWDAVEREDS AVVAEELGID AKALHMTLPA
 961 LSSWRRRERQ RRKVQRWRYR VEWKRLPNSR AQESLQGGWL LVVPQGRAGD VRVTQSVAEV
1021 AAKGGEATVL EVDALHPDRA AYAEALTRWP GVRGVVSFLA WEEQALAEHP VLSAGLAASL
1081 ALAQALIDVG GSGESAPRLW LVTEAAVVIG AADTGAVIDP VHAQLWGFGR VLALEHPELW
1141 GGLIDLPAVA GEPGSITDHA HADLLATVLA TMVQAAARGE DQVAVRTTGT YVPRLVRSGG
1201 SAHSGARRWQ PRDTVLVTGG MGPLTAHIVR WLADNGADQV VLLGGQGADG EAEALRAEFD
1261 GHTTKIELAD VDTEDSDALR SLLDRTTGEH PLRAVIHAPT VVEFASVAES DLVRFARTIS
1321 SKIAGVEQLD EVLSGIDTAH DVVFFSSVAG VWGSAGQSAY AAGNAFLDAV AQHRRLRGLP
1381 GTSVAWTPWD DDRSLASLGD SYLDRRGLRA LSIPGALASL QEVLDQDEVH AVVADVDWER
1441 FYAGFSAVRR TSFFDDVHDA HRPALSTAAT NDGQARDEDG GTELVRRLRP LTETEQQREL
1501 VSLVQSEVAA VLGHSSTDAV QPQRAFREIG FDSLTAVQLR NRLTATTGMR LPTTLVFDYP
1561 TTNGLAEYLR SELFGVSGAP ADLSVVRNAD EEDDPVVIVG MACRFPGGID TPEAFWKLLE
1621 AGGDVISELP ANRGWDMERL LNPDPEAKGT SATRYGGFLY DAGEFDAAFF GISPREALAM
1681 DPQQRLLLET VWELIESAGV APDSLHRSRT GTFIGSNGQF YAPLLWNSGG DLEGYQGVGN
1741 AGSVMSGRVA YSLGLEGPAV TVDTACSSSL VALHLAVQAL RRGECSLAIA GGVTVMSTPD
1801 SFVEFSRQQG LSEDGRCKAF ASTADGFGLA EGVSALLVER LSDARRLGHR VLAVVRGSAV
1861 NQDGASNGLT APNGPSQQRV IRAALADAGL APADVDVVEA HGTGTRLGDP IEAQALLATY
1921 GQGRAGGRPV VLGSVKSNIG HTQAAAGVAG VMKMVLALER GVVPKTLHVD EPSPHVDWSA
1981 GEVELAVEAV PWSRGGRVRR AGVSSFGISG TNAHVIVEEA PAEPEPEPGT RVVAAGDLVV
2041 PWVVSGRDAG ALREQAARLA AHVSSTGAGV VDVGWSLVAT RSVFEHRAVM VGSELDSMAE
2101 SLAGFAAGGV VPGVVSGVAP AEGRRVVFVF PGQGSQWVGM AAGLLDACPV FAEAVAECAA
2161 VLDPVTGWSL VEVLRGGGEA VLGRVDVVQP ALWAVMVSLA RTWRYYGVEP AAVVGHSQGE
2221 IAAACVAGGL SLADGARVVV LRSRAIARIA GGGGMVSVGL SAERVRTMLD TYGGRVSVAA
2281 VNGPSSTVVS GDVQALDELL AGCEREGVRA RRVPVDYASH SAQMDQLRDE LLEALADITP
2341 QHSSVPFFST VTADWLDTTA LDAGYWFTNL RETVRFQEAV EGLVAQGMGA FVECSPHPVL
2401 VPGIEQTLDA LDQNAAVLGS LRRDEGGLDR LLTSLAEAFV QGVPVDWTHA FEGMTPRTVD
2461 LPTYPFQRQH YWPKPAPAPG ANLGDVASVG LTAAGHPLLG AVVEMPDSDG LVLTGQISLR
2521 THPWLADHEV LGSVLLPGTA FVELAVQAAD RAGYDVLDEL TLEAPLVLPD RGGIQVRLAL
2581 GPSEADGRRS LQLHSRPEEA AGFHRWTRHA SGFVVPGGTG AARPTEPAGV WPPAGAEPVA
2641 LASDRYARLV ERGYTYGPSF QGLHTAWRHG DDVYAEVALP EGTPADGYAL HPALLDAAVQ
2701 AVGLGSFVED PGQVYLPFLW SDVTLHATGA TSLRVRVSPA GPDTVALALA DPAGAPVATV
2761 GALRLRTTSA AQLARARGSA EHAMFRVEWV EEGSAADRCR GGAGGTTYEG ERAAEAGAAA
2821 GTWAVLGPRV PAAVRTMGVD VVTALDTPDH PADPQSLADL AALGDTVPDV VVVTSLLSLA
2881 SGADSPLGNR PRPTAAEQDT AATVAGVHSA LHAALDLVQA WLADERHTAS RLVLVTRHAM
2941 TVAESDPEPD LLLAPVWGLV RSAQAENPGR FVLADIDGDE ASWDALPRAV ASAASEVAIR
3001 AGAVYVPRLA RATDEGLVVA DEAAGPWRLD VTEAGTLANL ALVPCPDASR PLGPDEVRIA
3061 VRAAGVNFRD VLLALGMYPD EGLMGAEAAG VVTEVGGGVT TLAPGDRVMG LVTGGFGPVA
3121 VTHHRMLVRM PRGWSFAEAA SVPVAFLTAY YALHDLAGLR GGESVLVHSA AGGVGMAAVQ
3181 LARHWDAEVF GTASKGKWDV LAAQGLDEEH IGSSRTTEFE QRFRATSGGR GIDVVLNALS
3241 GDFVDASARL LREGGRFVEM GKTDIRTDLG VVGADGVPDI RYVAFDLAEA GAERIGQMLD
3301 EIMALFDAGV LRLPPLRAWP VRRAHEALRF VSQARHVGKV VLTVPAALDA EGTVLITGAG
3361 TLGALVARHL VTEHDVRRLL LVSRSGVAPD LAAELGALGA EVTVAACDVA NRKALKALLE
3421 DIPPEHPVTG IVHTAGVLDD GVVSGLTPER VDTVLKPKVD AALTLESVIG ELDLDPALFV
3481 IFSSAASMLG GPGQGSYAAA NQFLDTLARH RARRGLTSVS LGWGLWHEAS GLTGGLADID
3541 RDRMSRAGIA PMPTDEALHL FDRATELGDP VLLPMRLNEA ALEDRAADGT LPPLLSGLVR
3601 VRHRPSARAG TATAAPATGP EAFARELAAA PDPRRALRDL VRGHVALVLG HSGPEAIDAE
```

```
3661 QAFRDIGFDS LTAVELRNRL NAETGLRLPG TLVFDYPNPS ALADHLLELL APATQPTAAP
3721 LLAELERVEQ LLSAAASPGG PASAVDEETR TLIATRLATL ASQWTHLPVG SPGNADNRSG
3781 PGESGQAQES GATGEHTAAW TSDDDLFAFL DKRLET
```

8,8a-deoxyoleandolide synthase 3:

```
   1 VAEAEKLREY LWRATTELKE VSDRLRETEE RAREPIAIVG MSCRFPGGGD ATVNTPEQFW
  61 DLLNSGGDGI AGLPEDRGWD LGRLYDPDPD RAGTSYVREG GFLYDSGEFD AAFFGISPRE
 121 ALAMDPQQRL LLETSWEAFE SAGIKRAALR GSDTGVYIGA WSTGYAGSPY RLVEGLEGQL
 181 AIGTTLGAAS GRVAYTFGLE GPAVTVDTAC SSSLVALHLA VQGLRRGECS LALVGGVTVM
 241 SSPVTLTTFS RQRGLSVDGR CKAFPASADG FGAAEGVGVL LVERLSDARR LGHRVLAVVR
 301 GSAVNQDGAS NGLTAPNGPS QQRVIRAALA DAGLAPADVD VVEAHGTGTR LGDPIEAQAL
 361 LATYGQGRAG GRPVWLGSVK SNIGHTQAAA GVAGVMKMVL ALGRGVVPKT LHVDEPSPHV
 421 DWSAGAVELL TEERPWEPEA ERLRRAGISA FGVSGTNAHV IVEEAPAEPE PEPGTRVVAA
 481 GDLVVPWVVS GRDARALRAQ AARLAAHVSG VSAVDVGWSL VATRSVFEHR AVAIGSELDS
 541 MAGSLAGFAA GGVVPGVVSG VAPAEGRRVV FVFPGQGSQW VGMAAGLLDA CPVFAEAVAE
 601 CAAVLDPVTG WSLVEVLQGR DATVLGRVDV VQPALWAVMV SLARTWRYYG VEPAAVVGHS
 661 QGEIAAACVA GGLSLADGAR VVVLRSRAIA RIAGGGGMVS VSLPAGRVRT MLEEFDGRLS
 721 VAAVNGPSST VVSGDVQALD ELLAGCEREG VRARRVPVDY ASHSAQMDQL RDELLEALAD
 781 ITPQDSSVPF FSTVTADWLG TTALGAGYWF TNLRETVRFQ EAVEGLVAQG MGAFVECSPH
 841 PVLVPGIEQT LDALDQNAAV FGSLRRDEGG LDRFLTSLAE AFVQGVPVDW SRAFEGVTPR
 901 TVDLPTYPFQ RQHYWLMAEE APVSQPPHSE NSFWSVVADA DAEAAAELLG VDVEAVEAVM
 961 PALSSWHRQS QLRAEVNQWR YDVAWKRLTT GALPEKPGNW LVVTPAGTDT TFAESLARTA
1021 AAELGVSVSF AQVDTAHPDR SQYAHALRQA LTGPENVDHL VSLLALDQAT DDLAAAPSCL
1081 AASLVLAQAL VDLGRVGEGP RLWLVTRGAV VAGPSDAGAV IDPVQAQVWG FGRVLGLEHP
1141 ELWGGLIDLP VGVDEEVCRR FVGVVASAGF EDQVAVRGSG VWVRRLVRAV VDGGGGGWRP
1201 RGTVLVTGGL GGLGAHTARW LVGGGADHVV LVSRRGGSAP GAGDLVRELE GLGGARVSVR
1261 ACDVADRVAL RALLSDLGEP VTAVFHAAGV PQSTPLAEIS VQEAADVMAA KVAGAVNLGE
1321 LVDPCGLEAF VLFSSNAGVW GSGGQAVYAA ANAFLDALAV RRRGVGLPAT SVAWGMWAGE
1381 GMASVGGAAR ELSRRGVRAM DPERAVAVMA DAVGRGEAFV AVADVDWERF VTGFASARPR
1441 PLISDLPEVR AVVEGQVQGR GQGLGLVGEE ESSGWLKRLS GLSRVRQEEE LVELVRAQAA
1501 VVLGHGSAQD VPAERAFKEL GFDSLTAVEL RNGLAAATGI RLPATMAFDH PNATAIARFL
1561 QSQLLPDAES ESAVPSSPED EVRQALASLS LDQLKGAGLL DPLLALTRLR EINSTVQNPE
1621 PTTESIDEMD GETCCAWRSA KSTAEPLTTG ADMPDPTAKY VEALRASLKE NERLRQQNHS
1681 LLAASREAIA ITAMSCRFGG GIDSPEDLWR FLAEGRDAVA GLPEDRGWDL DALYHPDPEN
1741 PGTTYVREGA FRYDAAQFDA GFFGISPREA LAMDPQQRLL LETSWELFER ADIDPYTVRG
1801 TATGIFIGAG HQGYGPDPKR APESVAGYLL TGTASAVLSG RISYTFGLEG PAVTVDTACS
1861 SSLVALHLAV QALRRGECSL AIAGGVAVMS TPDAFVEFSR QQGMARDGRC KAFAAAADGM
1921 GWGEGVSLLL LERLSDARRL GHRVLAVVRG SAVNQDGASN GLAAPNGPSQ QRVIRAALAD
1981 AGLAPADVDV VEAHGTGTRL GDPIEAQALL ATYGQGRAGG RPVWLGSVKS NIGHTQAAAG
2041 VAGVMKMVLA LGRGVVPKTL HVDEPSPHVD WSAGAVELLT EERPWEPEAE RLRRAGISAF
2101 GVSGTNAHVI VEEAPAEPEP EPGTRVVAAG DLVVPWVVSG RDVGALREQA ARLAAHVSST
2161 GAGVVDVGWS LVATRSVFEH RAVMVGTDLD SMAGSLAGFA AGGVVPGVVS GVAPAEGRRV
2221 VFVFPGQGSQ WVGMAAGLLD ACPVFAEAVA ECAAVLDPVT GWSLVEVLQG RDATVLGRVD
2281 VVQPALWAVM VSLARTWRYY GVEPAAVVGH SQGEIAAACV AGGLSLADGA RVVVLRSRAI
2341 ARIAGGGGMV SVSLPAGRVR TMLDTYGGRV SVAAVNGPSS TVVSGDVQAL DELLAGCERE
2401 GVRARRVPVD YASHSAQMDQ LRDELLEALA DITPQDSSVP FFSTVTADWL DTTALDAGYW
2461 FTNLRETVRF QEAVEGLVAQ GMGAFVECSP HPVLVPGIEQ TLDALDQNAA VLGSLRRDEG
2521 GLDRLLTSLA EAFVQGVPVD WTHAFEGVTP RTVDLPTYPF QRQRFWLDGS PASSANGVDG
2581 EADAMIWDAV EREDSVAVAE ELGIDAEALH TVLPALSSWR RRRVEHRRLQ DWRYRVEWKP
2641 FPAALDEVLG GGWLFVVPRG LADDGVVARV VAAVTARGGE VSVVELDPTR PDRRAYAEAV
2701 AGRGVSGVVS FLSWDDRRHS EHPVVPAGLA ASLVLAQALV DLGRVGEGPR LWLVTRDAVV
2761 AGPSDAGAVI DPVQAQVWGF GRVLGLEHPE LWGGLIDLPV EAPEPGSTCD HTYADLLATV
2821 VASAGFEDQV AVRGSGVWVR RLVRAVVDGG GGGWRPRGTV LVTGGLGGLG AHTARWLVGG
2881 GADHVVLVSR RGGSAPGAGD LVRELEGLGG ARVSVRACDV ADRVALRALL SDLGEPVTAV
2941 FHAAGVPQST PLAEISVQEA ADVMAAKVAG AVNLGELVDP CGLEAFVLFS SNAGVWGSGG
3001 QAVYAAANAF LDALAVRRRG VGLPATSVAW GMWAGEGMAS VGGAARELSR RGVRAMDPER
3061 AVAVMADAVG RGEAFVAVAD VDWERFVTGF ASARPRPLIS DLPEVRTALR NQEQEQLHAP
3121 VPEDRSAQLL RRLSMLSPAG REAELVKLVR TEAAAVLGHG SAQDVPAERA FKELGFDSLT
3181 AVQLRNRLAA ATGTRLPASA VFDHPHAAAL ARWLLAGMRH ADGGHGGGHA GGPGPDADEG
3241 RSAGAGHSGM LADLYRRSAE LGRSREFIGL LADTAAFRPV FHGPADLDAP LEAVPLADGV
3301 RKPQLICCSG TAPVGGPHEF ARLASFFRGT RAVSALPLPG YLPGEQLPAD LDAVLAAQAE
3361 AVEKQTGGAP FVLVGYSAGG LMAHALACHL AGRGTPPSGE VLVDVYPPGR QEPVFGWQKE
3421 LTEGMFAQDF VPMDDTRLTA LGTYDRLMGE WRPAPSGLPT LLIRATEPMA EWTGAIDWRA
3481 SWEYDHTAVD MPGNHFTIMR EHAEDAARHI DVWLKGLTP
```

The recombinant DNA compounds of the invention that encode the oleandolide PKS proteins or portions thereof are useful in a variety of applications. While many of these applications relate to the heterologous expression of the oleandolide PKS or the construction of hybrid PKS enzymes, many useful applications involve the natural oleandomycin producer *Streptomyces antibioticus*.

For example, one can use the recombinant DNA compounds of the invention to disrupt the *oleAI*, *oleAII*, or *oleAIII* genes by homologous recombination in *Streptomyces antibioticus*. The resulting host cell is a preferred host cell for making polyketides modified by oxidation, hydroxylation, and glycosylation in a manner similar to oleandomycin, because the genes that encode the proteins that perform these reactions are present in the host cell. Such a host cell also does not naturally produce any oleandomycin that could interfere with production or purification of the polyketide of interest.

One illustrative recombinant host cell provided by the present invention expresses a recombinant oleandolide PKS in which the module 1 KS domain is inactivated by deletion or other mutation. In a preferred embodiment, the inactivation is mediated by a change in the KS domain that renders it incapable of binding substrate (the KS1° mutation). In a particularly preferred embodiment, this inactivation is rendered by a mutation in the codon for the active site cysteine that changes the codon to another codon, such as an alanine codon. Such constructs are especially useful when placed in translational reading frame with extender modules 1 and 2 of an oleandolide or the corresponding modules of another PKS. The utility of these constructs is that host cells expressing, or cell free extracts containing, a PKS comprising the protein encoded thereby can be fed or supplied with N-acylcysteamine thioesters of precursor molecules to prepare a polyketide of interest. See U.S. patent application Serial No. 60/117,384, filed 27 Jan. 1999, and PCT patent publication No. US99/03986, both of which are incorporated herein by reference. Such KS1° constructs of the invention are useful in the production of 13-substituted-oleandomycin compounds in *Streptomyces antibioticus* host cells. Preferred compounds of the invention include those compounds in which the substituent at the 13-position is propyl, vinyl, propargyl, other lower alkyl, and substituted alkyl.

The compounds of the invention can also be used to construct recombinant host cells of the invention in which coding sequences for one or more domains or

modules of the oleandolide PKS have been deleted by homologous recombination with the *Streptomyces antibioticus* chromosomal DNA. Those of skill in the art will appreciate that such compounds are characterized by their homology with the chromosomal DNA and not by encoding a functional protein due to their intended function of deleting or otherwise altering portions of chromosomal DNA. For this and a variety of other applications, the compounds of the present invention include not only those DNA compounds that encode functional proteins but also those DNA compounds that are complementary or identical to any portion of the oleandolide PKS genes.

Thus, the invention provides a variety of modified *Streptomyces antibioticus* host cells in which one or more of the genes in the oleandolide PKS gene cluster have been mutated or disrupted. These cells are especially useful when it is desired to replace the disrupted function with a gene product expressed by a recombinant DNA expression vector. While such expression vectors of the invention are described in more detail in the following Section, those of skill in the art will appreciate that the vectors have application to *S. antibioticus* as well. Such *S. antibioticus* host cells can be preferred host cells for expressing oleandolide derivatives of the invention. Particularly preferred host cells of this type include those in which the coding sequence for the loading module has been mutated or disrupted, those in which one or more of any of the PKS gene ORFs has been mutated or disrupted, and/or those in which the genes for one or more oleandolide modification enzymes (glycosylation, epoxidation) have been mutated or disrupted.

While the present invention provides many useful compounds having application to, and recombinant host cells derived from, *Streptomyces antibioticus*, many important applications of the present invention relate to the heterologous expression of all or a portion of the oleandolide PKS genes in cells other than *S. antibioticus*, as described in the following Section.

Section II: Heterologous Expression of the Oleandolide PKS

In one important embodiment, the invention provides methods for the heterologous expression of one or more of the oleandolide PKS genes and recombinant DNA expression vectors useful in the method. For purposes of the invention, any host cell other than *Streptomyces antibioticus* is a heterologous host

cell. Thus, included within the scope of the invention in addition to isolated nucleic acids encoding domains, modules, or proteins of the oleandolide PKS, are recombinant expression vectors that include such nucleic acids. The term expression vector refers to a nucleic acid that can be introduced into a host cell or cell-free transcription and translation system. An expression vector can be maintained permanently or transiently in a cell, whether as part of the chromosomal or other DNA in the cell or in any cellular compartment, such as a replicating vector in the cytoplasm. An expression vector also comprises a promoter that drives expression of an RNA, which is translated into a polypeptide in the cell or cell extract. For efficient translation of RNA into protein, the expression vector also typically contains a ribosome-binding site sequence positioned upstream of the start codon of the coding sequence of the gene to be expressed. Other elements, such as enhancers, secretion signal sequences, transcription termination sequences, and one or more marker genes by which host cells containing the vector can be identified and/or selected, may also be present in an expression vector. Selectable markers, i.e., genes that confer antibiotic resistance or sensitivity, are preferred and confer a selectable phenotype on transformed cells when the cells are grown in an appropriate selective medium.

The various components of an expression vector can vary widely, depending on the intended use of the vector and especially the host cell(s) in which the vector is intended to replicate or drive expression. Expression vector components suitable for the expression of genes and maintenance of vectors in *E. coli*, yeast, *Streptomyces*, and other commonly used cells are widely known and commercially available. For example, suitable promoters for inclusion in the expression vectors of the invention include those that function in eucaryotic or procaryotic host cells. Promoters can comprise regulatory sequences that allow for regulation of expression relative to the growth of the host cell or that cause the expression of a gene to be turned on or off in response to a chemical or physical stimulus. For *E. coli* and certain other bacterial host cells, promoters derived from genes for biosynthetic enzymes, antibiotic-resistance conferring enzymes, and phage proteins can be used and include, for example, the galactose, lactose (*lac*), maltose, tryptophan (*trp*), beta-lactamase (*bla*), bacteriophage lambda PL, and T5 promoters. In addition, synthetic promoters, such as the *tac* promoter (U.S. Patent No. 4,551,433), can also be used.

Thus, recombinant expression vectors contain at least one expression system, which, in turn, is composed of at least a portion of the oleandolide PKS coding sequences operably linked to a promoter and optionally termination sequences that operate to effect expression of the coding sequence in compatible host cells. The host cells are modified by transformation with the recombinant DNA expression vectors of the invention to contain the expression system sequences either as extrachromosomal elements or integrated into the chromosome. The resulting host cells of the invention are useful in methods to produce PKS and post-PKS tailoring (modification) enzymes as well as polyketides and antibiotics and other useful compounds derived therefrom.

Preferred host cells for purposes of selecting vector components for expression vectors of the present invention include fungal host cells such as yeast and procaryotic host cells such as *E. coli* and *Streptomyces*, but mammalian cell cultures can also be used. In hosts such as yeasts, plants, or mammalian cells that ordinarily do not produce modular polyketide synthase enzymes, it may be necessary to provide, also typically by recombinant means, suitable holo-ACP synthases to convert the recombinantly produced PKS to functionality. Provision of such enzymes is described, for example, in PCT publication Nos. WO 97/13845 and 98/27203, each of which is incorporated herein by reference. Particularly preferred host cells for purposes of the present invention are *Streptomyces* and *Saccharopolyspora* host cells, as discussed in greater detail below.

In a preferred embodiment, the expression vectors of the invention are used to construct a heterologous recombinant *Streptomyces* host cell that expresses a recombinant PKS of the invention. *Streptomyces* is a convenient host for expressing polyketides, because polyketides are naturally produced in certain *Streptomyces* species, and *Streptomyces* cells generally produce the precursors needed to form the desired polyketide. Those of skill in the art will recognize that, if a *Streptomyces* host cell produces any portion of a PKS enzyme or produces a polyketide-modifying enzyme, the recombinant vector need drive expression of only those genes constituting the remainder of the desired PKS enzyme or other polyketide-modifying enzymes. Thus, such a vector may comprise only a single ORF, with the desired remainder of the polypeptides constituting the PKS provided by the genes on the host cell chromosomal DNA. If a *Streptomyces* or other host cell ordinarily produces polyketides, it may be desirable to modify the host so as to prevent the production of

endogenous polyketides prior to its use to express a recombinant PKS of the invention. Such modified hosts include *S. coelicolor* CH999 and similarly modified *S. lividans* described in U.S. Patent No. 5,672,491, and PCT publication Nos. WO 95/08548 and WO 96/40968, incorporated herein by reference. In such hosts, it may not be necessary to provide enzymatic activities for all of the desired post-translational modifications of the enzymes that make up the recombinantly produced PKS, because the host naturally expresses such enzymes. In particular, these hosts generally contain holo-ACP synthases that provide the pantotheinyl residue needed for functionality of the PKS.

The invention provides a wide variety of expression vectors for use in *Streptomyces*. The replicating expression vectors of the present invention include, for example and without limitation, those that comprise an origin of replication from a low copy number vector, such as SCP2* (see Hopwood *et al.*, *Genetic Manipulation of Streptomyces: A Laboratory manual* (The John Innes Foundation, Norwich, U.K., 1985); Lydiate *et al.*, 1985, *Gene 35*: 223-235; and Kieser and Melton, 1988, *Gene 65*: 83-91, each of which is incorporated herein by reference), SLP1.2 (Thompson *et al.*, 1982, *Gene 20*: 51-62, incorporated herein by reference), and pSG5(ts) (Muth *et al.*, 1989, *Mol. Gen. Genet. 219*: 341-348, and Bierman *et al.*, 1992, *Gene 116*: 43-49, each of which is incorporated herein by reference), or a high copy number vector, such as pIJ101 and pJV1 (see Katz *et al.*, 1983, *J. Gen. Microbiol. 129*: 2703-2714; Vara *et al.*, 1989, *J. Bacteriol. 171*: 5782-5781; and Servin-Gonzalez, 1993, *Plasmid 30*: 131-140, each of which is incorporated herein by reference). High copy number vectors are generally, however, not preferred for expression of large genes or multiple genes. For non-replicating and integrating vectors and generally for any vector, it is useful to include at least an *E. coli* origin of replication, such as from pUC, p1P, p1I, and pBR. For phage based vectors, the phage phiC31 and its derivative KC515 can be employed (see Hopwood *et al.*, *supra*). Also, plasmid pSET152, plasmid pSAM, plasmids pSE101 and pSE211, all of which integrate site-specifically in the chromosomal DNA of *S. lividans*, can be employed for purposes of the present invention.

The *Streptomyces* recombinant expression vectors of the invention typically comprise one or more selectable markers, including antibiotic resistance conferring genes selected from the group consisting of the *ermE* (confers resistance to

erythromycin and lincomycin), *tsr* (confers resistance to thiostrepton), *aadA* (confers resistance to spectinomycin and streptomycin), *aacC4* (confers resistance to apramycin, kanamycin, gentamicin, geneticin (G418), and neomycin), *hyg* (confers resistance to hygromycin), and *vph* (confers resistance to viomycin) resistance conferring genes. Alternatively, several polyketides are naturally colored, and this characteristic can provide a built-in marker for identifying cells.

Preferred *Streptomyces* host cell/vector combinations of the invention include *S. coelicolor* CH999 and *S. lividans* K4-114 and K4-155 host cells, which have been modified so as not to produce the polyketide actinorhodin, and expression vectors derived from the pRM1 and pRM5 vectors, as described in U.S. Patent No. 5,830,750 and U.S. patent application Serial Nos. 08/828,898, filed 31 Mar. 1997, and 09/181,833, filed 28 Oct. 1998, each of which is incorporated herein by reference. These vectors are particularly preferred in that they contain promoters compatible with numerous and diverse *Streptomyces spp.* Particularly useful promoters for *Streptomyces* host cells include those from PKS gene clusters that result in the production of polyketides as secondary metabolites, including promoters from aromatic (Type II) PKS gene clusters. Examples of Type II PKS gene cluster promoters are *act* gene promoters and *tcm* gene promoters; examples of Type I PKS gene cluster promoter are the spiramycin PKS and DEBS genes promoter. The present invention also provides the oleandolide PKS gene promoter in recombinant form. The promoter for the *oleA* genes is located upstream of the *oleAI* gene on cosmid pKOS055-5 of the invention. This promoter is contained within an ~1 kb segment upstream of the *oleAI* coding sequence and can be used to drive expression of the oleandolide PKS or any other coding sequence of interest in host cells in which the promoter functions, particularly *S. antibioticus* and generally any *Streptomyces* species.

As described above, particularly useful control sequences are those that alone or together with suitable regulatory systems activate expression during transition from growth to stationary phase in the vegetative mycelium. The promoter contained in the aforementioned plasmid pRM5, i.e., the *actI/actIII* promoter pair and the *actII-ORF4* activator gene, is particularly preferred. Other useful *Streptomyces* promoters include without limitation those from the *ermE* gene and the *melC1* gene, which act constitutively, and the *tipA* gene and the *merA* gene, which can be induced at any

growth stage. In addition, the T7 RNA polymerase system has been transferred to *Streptomyces* and can be employed in the vectors and host cells of the invention. In this system, the coding sequence for the T7 RNA polymerase is inserted into a neutral site of the chromosome or in a vector under the control of the inducible *merA* promoter, and the gene of interest is placed under the control of the T7 promoter. As noted above, one or more activator genes can also be employed to activate initiation of transcription at promoter sequences. Activator genes in addition to the *actII-ORF4* gene described above include *dnrI*, *redD*, and *ptpA* genes (see U.S. patent application Serial No. 09/181,833, supra).

To provide a preferred host cell and vector for purposes of the invention, the oleandolide PKS genes were placed on a recombinant expression vector that was transferred to the non-macrolide producing host *Streptomyces lividans* K4-114, as described in Example 4. Transformation of *S. lividans* K4-114 (strain K4-155 can also be used) with this expression vector resulted in a strain which produced detectable amounts of 8,8a-deoxyoleandolide as determined by analysis of extracts by LC/MS.

Moreover, and as noted in the preceding Section, the present invention also provides recombinant DNA compounds in which the encoded oleandolide module 1 KS domain is inactivated or absent altogether. Example 4 below describes the introduction into *Streptomyces lividans* of a recombinant expression vector of the invention that encodes an oleandolide PKS with a KS1° domain. The resulting host cells can be fed or supplied with N-acylcysteamine thioesters of precursor molecules to prepare oleandolide derivatives. Such cells of the invention are especially useful in the production of 13-substituted-6-deoxyerythronolide B compounds in recombinant host cells. Preferred compounds of the invention include those compounds in which the substituent at the 13-position is propyl, vinyl, propargyl, other lower alkyl, and substituted alkyl. The unmodified polyketides, called macrolide aglycones, produced in *S. lividans* K4-114 or K4-155 can be hydroxylated and glycosylated by adding them to the fermentation of a strain, such as, for example, *S. antibioticus* or *Saccharopolyspora erythraea*, that contains the requisite modification enzymes.

There are a wide variety of diverse organisms that can modify macrolide aglycones to provide compounds with, or that can be readily modified to have, useful activities. For example, *Saccharopolyspora erythraea* can convert 6-dEB and oleandolide to a variety of useful compounds. The erythronolide 6-dEB is converted

by the *eryF* gene product to erythronolide B, which is, in turn, glycosylated by the *eryB* gene product to obtain 3-O-mycarosylerythronolide B, which contains L-mycarose at C-3. The enzyme *eryC* gene product then converts this compound to erythromycin D by glycosylation with D-desosamine at C-5. Erythromycin D, therefore, differs from 6-dEB through glycosylation and by the addition of a hydroxyl group at C-6. Erythromycin D can be converted to erythromycin B in a reaction catalyzed by the *eryG* gene product by methylating the L-mycarose residue at C-3. Erythromcyin D is converted to erythromycin C by the addition of a hydroxyl group at C-12 in a reaction catalyzed by the *eryK* gene product. Erythromycin A is obtained from erythromycin C by methylation of the mycarose residue in a reaction catalyzed by the *eryG* gene product.

The unmodified oleandolide compounds provided by the present invention, such as, for example, the oleandolide produced in *Streptomyces lividans*, can be provided to cultures of *Saccharopolyspora erythraea* and converted to the corresponding derivatives of erythromycins A, B, C, and D in accordance with the procedure provided in Example 6, below. To ensure that only the desired compound is produced, one can use an *S. erythraea eryA* mutant that is unable to produce 6-dEB but can still carry out the desired conversions (Weber *et al.*, 1985, *J. Bacteriol. 164*(1): 425-433). Also, one can employ other mutant strains, such as *eryB*, *eryC*, *eryG*, and/or *eryK* mutants, or mutant strains having mutations in multiple genes, to accumulate a preferred compound. The conversion can also be carried out in large fermentors for commercial production.

Moreover, there are other useful organisms that can be employed to hydroxylate and/or glycosylate the compounds of the invention. As described above, the organisms can be mutants unable to produce the polyketide normally produced in that organism, the fermentation can be carried out on plates or in large fermentors, and the compounds produced can be chemically altered after fermentation. Thus, *Streptomyces venezuelae*, which produces picromycin, contains enzymes that can transfer a desosaminyl group to the C-5 hydroxyl and a hydroxyl group to the C-12 position. In addition, *S. venezuelae* contains a glucosylation activity that glucosylates the 2'-hydroxyl group of the desosamine sugar. This latter modification reduces antibiotic activity, but the glucosyl residue is removed by cellular enzymatic action. Another organism, *S. narbonensis*, contains the same modification enzymes as *S.*

venezuelae, except the C-12 hydroxylase. Thus, the present invention provides the compounds produced by hydroxylation and glycosylation of the macrolide aglycones of the invention by action of the enzymes endogenous to *S. narbonensis* and *S. venezuelae*.

Other organisms suitable for making compounds of the invention include *Streptomyces antibioticus* (discussed in the preceding Section), *Micromonospora megalomicea*, *S. fradiae*, and *S. thermotolerans*. *M. megalomicea* produces megalomicin and contains enzymes that hydroxylate the C-6 and C-12 positions and glycosylate the C-3 hydroxyl with mycarose, the C-5 hydroxyl with desosamine, and the C-6 hydroxyl with megosamine (also known as rhodosamine), as well as acylating various positions. In addition to antibiotic activity, compounds of the invention produced by treatment with *M. megalomicea* enzymes can have antiparasitic activity as well. *S. fradiae* contains enzymes that glycosylate the C-5 hydroxyl with mycaminose and then the 4'-hydroxyl of mycaminose with mycarose, forming a disaccharide. *S. thermotolerans* contains the same activities as well as acylation activities. Thus, the present invention provides the compounds produced by hydroxylation and glycosylation of the macrolide aglycones of the invention by action of the enzymes endogenous to *S. antibioticus*, *M. megalomicea*, *S. fradiae*, and *S. thermotolerans*.

The present invention also provides methods and genetic constructs for producing the glycosylated and/or hydroxylated compounds of the invention directly in the host cell of interest. Thus, the recombinant genes of the invention, which include recombinant *oleAI*, *oleAII*, and *oleAIII* genes with one or more deletions and/or insertions, including replacements of an *oleA* gene fragment with a gene fragment from a heterologous PKS gene (as discussed in the next Section), can be included on expression vectors suitable for expression of the encoded gene products in *Saccharopolyspora erythraea*, *Streptomyces antibioticus*, *S. venezuelae*, *S. narbonensis*, *Micromonospora megalomicea*, *S. fradiae*, and *S. thermotolerans*. A number of erythromycin high-producing strains of *S. erythraea* have been developed, and in a preferred embodiment, the oleandolide PKS genes are introduced into such strains (or erythromycin non-producing mutants thereof) to provide the corresponding modified oleandolide compounds in high yields.

Moreover, additional recombinant gene products can be expressed in the host cell to improve production of a desired polyketide. As but one non-limiting example, certain recombinant PKS proteins of the invention may produce a polyketide other than or in addition to the predicted polyketide, because the polyketide is cleaved from the PKS by the thioesterase (TE) domain in module 6 prior to processing by other domains on the PKS, in particular, any KR, DH, and/or ER domains in module 6. The production of the predicted polyketide can be increased in such instances by deleting the TE domain coding sequences from the gene and, optionally, expressing the TE domain as a separate protein. See Gokhale *et al.*, Feb. 1999, "Mechanism and specificity of the terminal thioesterase domain from the erythromycin polyketide synthase," *Chem. & Biol. 6*: 117-125, incorporated herein by reference.

Thus, in one important aspect, the present invention provides methods, expression vectors, and recombinant host cells that enable the production of oleandolide and hydroxylated and glycosylated derivatives of oleandolide in heterologous host cells. The present invention also provides methods for making a wide variety of polyketides derived in part from the oleandolide PKS, as described in the following Section.

Section III: Hybrid PKS Genes

The present invention provides recombinant DNA compounds encoding each of the domains of each of the modules of the oleandolide PKS. The availability of these compounds permits their use in recombinant procedures for production of desired portions of the oleandolide PKS fused to or expressed in conjunction with all or a portion of a heterologous PKS. The resulting hybrid PKS can then be expressed in a host cell to produce a desired polyketide.

Thus, in accordance with the methods of the invention, a portion of the oleandolide PKS coding sequence that encodes a particular activity can be isolated and manipulated, for example, to replace the corresponding region in a different modular PKS. In addition, coding sequences for individual modules of the PKS can be ligated into suitable expression systems and used to produce the portion of the protein encoded. The resulting protein can be isolated and purified or can may be employed *in situ* to effect polyketide synthesis. Depending on the host for the recombinant production of the domain, module, protein, or combination of proteins,

suitable control sequences such as promoters, termination sequences, enhancers, and the like are ligated to the nucleotide sequence encoding the desired protein in the construction of the expression vector, as described in the preceding Section.

In one important embodiment, the invention thus provides hybrid PKS enzymes and the corresponding recombinant DNA compounds that encode those hybrid PKS enzymes. For purposes of the invention, a hybrid PKS is a recombinant PKS that comprises all or part of one or more extender modules, loading module, and/or thioesterase/cyclase domain of a first PKS and all or part of one or more extender modules, loading module, and/or thioesterase/cyclase domain of a second PKS. In one preferred embodiment, the first PKS is most but not all of the oleandolide PKS, and the second PKS is only a portion or all of a non-oleandolide PKS. An illustrative example of such a hybrid PKS includes an oleandolide PKS in which the oleandolide PKS loading module has been replaced with a loading module of another PKS. Another example of such a hybrid PKS is an oleandolide PKS in which the AT domain of extender module 3 is replaced with an AT domain that binds only malonyl CoA. In another preferred embodiment, the first PKS is most but not all of a non-oleandolide PKS, and the second PKS is only a portion or all of the oleandolide PKS. An illustrative example of such a hybrid PKS includes a rapamycin PKS in which an AT specific for malonyl CoA is replaced with the AT from the oleandolide PKS specific for methylmalonyl CoA. Other illustrative hybrid PKSs of the invention are described below.

Those of skill in the art will recognize that all or part of either the first or second PKS in a hybrid PKS of the invention need not be isolated from a naturally occurring source. For example, only a small portion of an AT domain determines its specificity. See PCT patent application No. WO US99/15047, and Lau *et al.*, *infra*, incorporated herein by reference. The state of the art in DNA synthesis allows the artisan to construct *de novo* DNA compounds of size sufficient to construct a useful portion of a PKS module or domain. Thus, the desired derivative coding sequences can be synthesized using standard solid phase synthesis methods such as those described by Jaye *et al.*, 1984, *J. Biol. Chem. 259*: 6331, and instruments for automated synthesis are available commercially from, for example, Applied Biosystems, Inc. For purposes of the invention, such synthetic DNA compounds are deemed to be a portion of a PKS.

With this general background regarding hybrid PKSs of the invention, one can better appreciate the benefit provided by the DNA compounds of the invention that encode the individual domains, modules, and proteins that comprise the oleandolide PKS. As described above, the oleandolide PKS is comprised of a loading module, six extender modules composed of a KS, AT, ACP, and KR, DH, and ER domains, and a thioesterase domain. The DNA compounds of the invention that encode these domains individually or in combination are useful in the construction of the hybrid PKS encoding DNA compounds of the invention.

The recombinant DNA compounds of the invention that encode the loading module of the oleandolide PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the oleandolide PKS loading module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS protein or portion thereof. The resulting construct, in which the coding sequence for the loading module of the heterologous PKS is replaced by that for the coding sequence of the oleandolide PKS loading module provides a novel PKS. Examples include the 6-deoxyerythronolide B, rapamycin, FK-506, FK-520, rifamycin, and avermectin PKS protein coding sequences. In another embodiment, a DNA compound comprising a sequence that encodes the oleandolide PKS loading module is inserted into a DNA compound that comprises the coding sequence for the oleandolide PKS or a recombinant oleandolide PKS that produces an oleandolide derivative.

In another embodiment, a portion of the loading module coding sequence is utilized in conjuction with a heterologous coding sequence. In this embodiment, the invention provides, for example, replacing the malonyl CoA (acetyl CoA) specific AT with a propionyl CoA (methylmalonyl), butyryl CoA (ethylmalonyl), or other CoA specific AT. In addition, the KS^Q and/or ACP can be replaced by another inactivated KS and/or another ACP. Alternatively, the KS^Q and AT of the loading module can be replaced by an AT of a loading module such as that of DEBS. The resulting heterologous loading module coding sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes oleandolide, an oleandolide derivative, or another polyketide.

The recombinant DNA compounds of the invention that encode the first extender module of the oleandolide PKS and the corresponding polypeptides encoded

thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the oleandolide PKS first extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the first extender module of the oleandolide PKS or the latter is merely added to coding sequences for modules of the heterologous PKS, provides a novel PKS coding sequence. In another embodiment, a DNA compound comprising a sequence that encodes the first extender module of the oleandolide PKS is inserted into a DNA compound that comprises coding sequences for the oleandolide PKS or a recombinant oleandolide PKS that produces an oleandolide derivative.

In another embodiment, a portion or all of the first extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting (which includes inactivating) the KR; inserting a DH or a DH and ER; and/or replacing the KR with another KR, a DH and KR, or a DH, KR, and ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the oleandolide PKS, from a gene for a PKS that produces a polyketide other than oleandolide, or from chemical synthesis. The resulting heterologous first extender module coding sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes oleandolide, an oleandolide derivative, or another polyketide.

Those of skill in the art will recognize, however, that deletion of the KR domain of module 1 or insertion of a DH domain or DH and KR domains into module 1 will prevent the typical cyclization of the polyketide at the hydroxyl group created by the KR if such hybrid module is employed as a first extender module in a hybrid PKS or is otherwise involved in producing a portion of the polyketide at which cyclization is to occur. Such deletions or insertions can be useful, however, to create linear molecules or to induce cyclization at another site in the molecule.

As noted above, the invention also provides recombinant PKSs and recombinant DNA compounds and vectors that encode a PKS protein in which the KS domain of the first extender module has been inactivated. Such constructs are especially useful when placed in translational reading frame with the remaining modules and domains of an oleandolide or oleandolide derivative PKS, a hybrid PKS, or a heterologous PKS. The utility of these constructs is that host cells expressing, or cell free extracts containing, the PKS encoded thereby can be fed or supplied with N-acylcysteamine thioesters of precursor molecules to prepare oleandolide derivative compounds. See U.S. patent application Serial No. 60/117,384, filed 27 Jan. 1999, and PCT publication Nos. WO 99/03986 and 97/02358, each of which is incorporated herein by reference.

The recombinant DNA compounds of the invention that encode the second extender module of the oleandolide PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the oleandolide PKS second extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the second extender module of the oleandolide PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS. In another embodiment, a DNA compound comprising a sequence that encodes the second extender module of the oleandolide PKS is inserted into a DNA compound that comprises the coding sequences for the oleandolide PKS or a recombinant oleandolide PKS that produces an oleandolide derivative.

In another embodiment, a portion or all of the second extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting (or inactivating) the KR; replacing the KR with a KR, a KR and a DH, or a KR, DH, and ER; and/or inserting a DH or a DH and an ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another

module of the oleandolide PKS, from a coding sequence for a PKS that produces a polyketide other than oleandolide, or from chemical synthesis. The resulting heterologous second extender module coding sequence can be utilized in conjunction with a coding sequence from a PKS that synthesizes oleandolide, an oleandolide derivative, or another polyketide.

The recombinant DNA compounds of the invention that encode the third extender module of the oleandolide PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the oleandolide PKS third extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the third extender module of the oleandolide PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS. In another embodiment, a DNA compound comprising a sequence that encodes the third extender module of the oleandolide PKS is inserted into a DNA compound that comprises coding sequences for the oleandolide PKS or a recombinant oleandolide PKS that produces an oleandolide derivative.

In another embodiment, a portion or all of the third extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting the inactive KR; and/or replacing the KR with an active KR, or a KR and DH, or a KR, DH, and ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the oleandolide PKS, from a gene for a PKS that produces a polyketide other than oleandolide, or from chemical synthesis. The resulting heterologous third extender module coding sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes oleandolide, an oleandolide derivative, or another polyketide.

The recombinant DNA compounds of the invention that encode the fourth extender module of the oleandolide PKS and the corresponding polypeptides encoded

thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the oleandolide PKS fourth extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the fourth extender module of the oleandolide PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS. In another embodiment, a DNA compound comprising a sequence that encodes the fourth extender module of the oleandolide PKS is inserted into a DNA compound that comprises coding sequences for the oleandolide PKS or a recombinant oleandolide PKS that produces an oleandolide derivative.

In another embodiment, a portion of the fourth extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting or inactivating any one, two, or all three of the ER, DH, and KR; and/or replacing any one, two, or all three of the ER, DH, and KR with either a KR, a DH and KR, or a KR, DH, and ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the oleandolide PKS (except for the DH and ER domains), from a coding sequence for a PKS that produces a polyketide other than oleandolide, or from chemical synthesis. The resulting heterologous fourth extender module coding sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes oleandolide, an oleandolide derivative, or another polyketide.

The recombinant DNA compounds of the invention that encode the fifth extender module of the oleandolide PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the oleandolide PKS fifth extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the fifth extender module of the

oleandolide PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS. In another embodiment, a DNA compound comprising a sequence that encodes the fifth extender module of the oleandolide PKS is inserted into a DNA compound that comprises the coding sequence for the oleandolide PKS or a recombinant oleandolide PKS that produces an oleandolide derivative.

In another embodiment, a portion or all of the fifth extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting (or inactivating) the KR; inserting a DH or a DH and ER; and/or replacing the KR with another KR, a DH and KR, or a DH, KR, and ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the oleandolide PKS, from a coding sequence for a PKS that produces a polyketide other than oleandolide, or from chemical synthesis. The resulting heterologous fifth extender module coding sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes oleandolide, an oleandolide derivative, or another polyketide.

The recombinant DNA compounds of the invention that encode the sixth extender module of the oleandolide PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the oleandolide PKS sixth extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the sixth extender module of the oleandolide PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS. In another embodiment, a DNA compound comprising a sequence that encodes the sixth extender module of the oleandolide PKS is inserted into a DNA compound that comprises the coding sequences for the oleandolide PKS or a recombinant oleandolide PKS that produces an oleandolide derivative.

In another embodiment, a portion or all of the sixth extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting or inactivating the KR or replacing the KR with another KR, a KR and DH, or a KR, DH, and an ER; and/or inserting a DH or a DH and ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the oleandolide PKS, from a coding sequence for a PKS that produces a polyketide other than oleandolide, or from chemical synthesis. The resulting heterologous sixth extender module coding sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes oleandolide, an oleandolide derivative, or another polyketide.

The sixth extender module of the oleandolide PKS is followed by a thioesterase domain. This domain is important in the cyclization of the polyketide and its cleavage from the PKS. The present invention provides recombinant DNA compounds that encode hybrid PKS enzymes in which the oleandolide PKS is fused to a heterologous thioesterase or a heterologous PKS is fused to the oleandolide synthase thioesterase. Thus, for example, a thioesterase domain coding sequence from another PKS gene can be inserted at the end of the sixth (or other final) extender module coding sequence in recombinant DNA compounds of the invention or the oleandolide PKS thioesterase can be similarly fused to a heterologous PKS. Recombinant DNA compounds encoding this thioesterase domain are useful in constructing DNA compounds that encode the oleandolide PKS, a PKS that produces an oleandolide derivative, and a PKS that produces a polyketide other than oleandolide or an oleandolide derivative.

Thus, the hybrid modules of the invention are incorporated into a PKS to provide a hybrid PKS of the invention. A hybrid PKS of the invention can result not only:

(i) from fusions of heterologous domain (where heterologous means the domains in that module are from at least two different naturally occurring modules)

coding sequences to produce a hybrid module coding sequence contained in a PKS gene whose product is incorporated into a PKS,

but also:

(ii) from fusions of heterologous module (where heterologous module means two modules are adjacent to one another that are not adjacent to one another in naturally occurring PKS enzymes) coding sequences to produce a hybrid coding sequence contained in a PKS gene whose product is incorporated into a PKS,

(iii) from expression of one or more oleandolide PKS genes with one or more non-oleandolide PKS genes, including both naturally occurring and recombinant non-oleandolide PKS genes, and

(iv) from combinations of the foregoing.

Various hybrid PKSs of the invention illustrating these various alternatives are described herein.

An example of a hybrid PKS comprising fused modules results from fusion of the loading module of either DEBS or the narbonolide PKS (see PCT patent application No. US99/11814, incorporated herein by reference) with extender modules 1 and 2 of the oleandolide PKS to produce a hybrid *oleAI* gene. Co-expression of either one of these two hybrid *oleAI* genes with the *oleAII* and *oleAIII* genes in suitable host cells, such as *Streptomcyes lividans*, results in expression of a hybrid PKS of the invention that produces 6-deoxyerythronolide B in recombinant host cells. Co-expression of either one of these two hybrid *oleAI* genes with the *eryAII* and *eryAIII* genes similarly results in the production of 6-dEB, while co-expression with the analogous narbonolide PKS genes (*picAII* and *picAIII*) results in the production of 3-keto-6-dEB.

Another example of a hybrid PKS comprising a hybrid module is prepared by co-expressing the *oleAI* and *oleAII* genes with an *oleAIII* hybrid gene encoding extender module 5 and the KS and AT of extender module 6 of the oleandolide PKS fused to the ACP of extender module 6 and the TE of the narbonolide PKS. The resulting hybrid PKS of the invention produces 3-deoxy-3-oxo-8,8a-deoxyoleandolide (3-keto-oleandolide). This compound is useful in the production of 14-desmethyl ketolides, compounds with potent anti-bacterial activity. This compound can also be prepared by a recombinant oleandolide derivative PKS of the invention in which the KR domain of module 6 of the oleandolide PKS has been deleted or replaced with an

inactive KR domain. Moreover, the invention provides hybrid PKSs in which not only the above changes have been made but also the AT domain of module 6 has been replaced with a malonyl-specific AT. These hybrid PKSs produce 2-desmethyl-3-deoxy-3-oxo-8,8a-deoxyoleandolide, a useful intermediate in the preparation of 2,14-didesmethyl ketolides, compounds with potent antibiotic activity.

Another illustrative example of a hybrid PKS includes the hybrid PKS of the invention resulting only from the latter change in the hybrid PKS just described. Thus, co-expression of the *oleAI* and *oleAII* genes with a hybrid *oleAIII* gene in which the AT domain of module 6 has been replaced by a malonyl-specific AT results in the expression of a hybrid PKS that produces 2-desmethyl-8,8a-deoxyoleandolide in recombinant host cells. This compound is a useful intermediate for making 2,14-didesmethyl erythromycins in recombinant host cells of the invention.

While many of the hybrid PKSs described above are composed primarily of oleandolide PKS proteins, those of skill in the art recognize that the present invention provides many different hybrid PKSs, including those composed of only a small portion of the oleandolide PKS. For example, the present invention provides a hybrid PKS in which a hybrid *oleAI* gene that encodes the oleandolide loading module fused to extender modules 1 and 2 of DEBS is coexpressed with the *eryAII* and *eryAIII* genes. The resulting hybrid PKS produces 8,8a-deoxyoleandolide. When the construct is expressed in *Saccharopolyspora erythraea* host cells (either via chromosomal integration in the chromosome or via a vector that encodes the hybrid PKS), the resulting recombinant host cell of the invention produces 14-desmethyl erythromycins. Another illustrative example is the hybrid PKS of the invention composed of the *oleAI* and *eryAII* and *eryAIII* gene products. This construct is also useful in expressing 14-desmethyl erythromycins in *Saccharopolyspora erythraea* host cells, as described in Example 3, below. In a preferred embodiment, the *S. erythraea* host cells are *eryAI* mutants that do not produce 6-deoxyerythronolide B.

Another example is the hybrid PKS of the invention composed of the products of the *picAI* and *picAII* genes (the two proteins that comprise the loading module and extender modules 1 - 4, inclusive, of the narbonolide PKS) and the *oleAIII* gene. The resulting hybrid PKS produces the macrolide aglycone 3-hydroxy-narbonolide in *Streptomyces lividans* host cells and the corresponding erythromycins in

Saccharopolyspora erythraea host cells. This hybrid PKS of the invention is described in Example 5, below.

Each of the foregoing hybrid PKS enzymes of the invention, and the hybrid PKS enzymes of the invention generally, can be expressed in a host cell that also expresses a functional *oleP* gene product. Such expression provides the compounds of the invention in which the C-8-C-8a epoxide is present.

The following Table lists references describing illustrative PKS genes and corresponding enzymes that can be utilized in the construction of the recombinant hybrid PKSs and the corresponding DNA compounds that encode them of the invention. Also presented are various references describing tailoring enzymes and corresponding genes that can be employed in accordance with the methods of the invention.

Avermectin

U.S. Pat. No. 5,252,474 to Merck.

MacNeil *et al.*, 1993, Industrial Microorganisms: Basic and Applied Molecular Genetics, Baltz, Hegeman, & Skatrud, eds. (ASM), pp. 245-256, A Comparison of the Genes Encoding the Polyketide Synthases for Avermectin, Erythromycin, and Nemadectin.

MacNeil *et al.*, 1992, *Gene 115*: 119-125, Complex Organization of the *Streptomyces avermitilis* genes encoding the avermectin polyketide synthase.

Candicidin (FR008)

Hu *et al.*, 1994, *Mol. Microbiol.* 14: 163-172.

Epothilone

U.S. patent application Serial No. 60/130,560, filed 22 Apr. 1999, and Serial No. 60/122,620, filed 3 Mar. 1999.

Erythromycin

PCT Pub. No. 93/13663 to Abbott.

US Pat. No. 5,824,513 to Abbott.

Donadio *et al.*, 1991, *Science 252*:675-9.

Cortes *et al.*, 8 Nov. 1990, *Nature 348*:176-8, An unusually large multifunctional polypeptide in the erythromycin producing polyketide synthase of *Saccharopolyspora erythraea*.

Glycosylation Enzymes

PCT Pat. App. Pub. No. 97/23630 to Abbott.

FK-506

Motamedi *et al.*, 1998, The biosynthetic gene cluster for the macrolactone ring of the immunosuppressant FK506, *Eur. J. biochem. 256*: 528-534.

Motamedi *et al.*, 1997, Structural organization of a multifunctional polyketide synthase involved in the biosynthesis of the macrolide immunosuppressant FK506, *Eur. J. Biochem. 244*: 74-80.

Methyltransferase

US 5,264,355, issued 23 Nov. 1993, Methylating enzyme from *Streptomyces* MA6858. 31-O-desmethyl-FK506 methyltransferase.

Motamedi *et al.*, 1996, Characterization of methyltransferase and hydroxylase genes involved in the biosynthesis of the immunosuppressants FK506 and FK520, *J. Bacteriol. 178*: 5243-5248.

FK-520

U.S. patent application Serial No. 60/139,650, filed 17 Jun. 1999, and 60/123,810, filed 11 Mar. 1999. See also Nielsen *et al.*, 1991, *Biochem. 30*:5789-96 (enzymology of pipecolate incorporation).

Lovastatin

U.S. Pat. No. 5,744,350 to Merck.

Narbomycin (and Picromycin)

PCT patent application No. WO US99/11814, filed 28 May 1999.

Nemadectin

MacNeil *et al.*, 1993, *supra.*

Niddamycin

Kakavas *et al.*, 1997, Identification and characterization of the niddamycin polyketide synthase genes from *Streptomyces caelestis*, *J. Bacteriol. 179*: 7515-7522.

Platenolide

EP Pat. App. Pub. No. 791,656 to Lilly.

Rapamycin

Schwecke *et al.*, Aug. 1995, The biosynthetic gene cluster for the polyketide rapamycin, *Proc. Natl. Acad. Sci. USA 92*:7839-7843.

Aparicio *et al.*, 1996, Organization of the biosynthetic gene cluster for rapamycin in *Streptomyces hygroscopicus*: analysis of the enzymatic domains in the modular polyketide synthase, *Gene 169*: 9-16.

Rifamycin

August *et al.*, 13 Feb. 1998, Biosynthesis of the ansamycin antibiotic rifamycin: deductions from the molecular analysis of the *rif* biosynthetic gene cluster of *Amycolatopsis mediterranei* S669, *Chemistry & Biology, 5*(2): 69-79.

Soraphen

U.S. Pat. No. 5,716,849 to Novartis.

Schupp *et al.*, 1995, *J. Bacteriology 177*: 3673-3679. A *Sorangium cellulosum* (Myxobacterium) Gene Cluster for the Biosynthesis of the Macrolide Antibiotic Soraphen A: Cloning, Characterization, and Homology to Polyketide Synthase Genes from Actinomycetes.

Spiramycin

U.S. Pat. No. 5,098,837 to Lilly.

Activator Gene

U.S. Pat. No. 5,514,544 to Lilly.

Tylosin

EP Pub. No. 791,655 to Lilly.

Kuhstoss *et al.*, 1996, *Gene 183*:231-6., Production of a novel polyketide through the construction of a hybrid polyketide synthase.

U.S. Pat. No. 5,876,991 to Lilly.

Tailoring enzymes

Merson-Davies and Cundliffe, 1994, *Mol. Microbiol. 13*: 349-355. Analysis of five tylosin biosynthetic genes from the *tylBA* region of the *Streptomyces fradiae* genome.

As the above Table illustrates, there are a wide variety of PKS genes that serve as readily available sources of DNA and sequence information for use in constructing the hybrid PKS-encoding DNA compounds of the invention. Methods for constructing hybrid PKS-encoding DNA compounds are described without reference to the oleandolide PKS in U.S. Patent Nos. 5,672,491 and 5,712,146 and PCT publication No. 98/49315, each of which is incorporated herein by reference.

In constructing hybrid PKSs of the invention, certain general methods may be helpful. For example, it is often beneficial to retain the framework of the module to be

altered to make the hybrid PKS. Thus, if one desires to add DH and ER functionalities to a module, it is often preferred to replace the KR domain of the original module with a KR, DH, and ER domain-containing segment from another module, instead of merely inserting DH and ER domains. One can alter the stereochemical specificity of a module by replacement of the KS domain with a KS domain from a module that specifies a different stereochemistry. See Lau *et al.*, 1999, "Dissecting the role of acyltransferase domains of modular polyketide synthases in the choice and stereochemical fate of extender units" *Biochemistry 38*(5):1643-1651, incorporated herein by reference. One can alter the specificity of an AT domain by changing only a small segment of the domain. See Lau *et al.*, *supra*. One can also take advantage of known linker regions in PKS proteins to link modules from two different PKSs to create a hybrid PKS. See Gokhale *et al.*, 16 Apr. 1999, "Dissecting and Exploiting Intermodular Communication in Polyketide Synthases", *Science 284*: 482-485, incorporated herein by reference.

The hybrid PKS-encoding DNA compounds of the invention can be and often are hybrids of more than two PKS genes. Even where only two genes are used, there are often two or more modules in the hybrid gene in which all or part of the module is derived from a second (or third) PKS gene. Thus, as one illustrative example, the invention provides a hybrid PKS that contains the naturally occurring loading module and thioesterase domain as well as extender modules one, two, four, and six of the oleandolide PKS and further contains hybrid or heterologous extender modules three and five. Hybrid or heterologous extender modules three and five contain AT domains specific for malonyl CoA and derived from, for example, the rapamycin PKS genes.

To construct a hybrid PKS or oleandolide PKS of the invention, one can employ a technique, described in PCT Pub. No. 98/27203 and U.S. provisional patent application Serial No. 60/129,731, filed 16 Apr. 99, incorporated herein by reference, in which the large oleandolide PKS gene cluster is divided into two or more, typically three, segments, and each segment is placed on a separate expression vector. In this manner, each of the segments of the gene can be altered, and various altered segments can be combined in a single host cell to provide a recombinant PKS gene of the invention. This technique makes more efficient the construction of large libraries of recombinant PKS genes, vectors for expressing those genes, and host cells comprising those vectors.

The invention also provides libraries of PKS genes, PKS proteins, and ultimately, of polyketides, that are constructed by generating modifications in the oleandolide PKS so that the protein complexes produced have altered activities in one or more respects and thus produce polyketides other than the oleandolide natural product of the PKS. Novel polyketides may thus be prepared, or polyketides in general prepared more readily, using this method. By providing a large number of different genes or gene clusters derived from a naturally occurring PKS gene cluster, each of which has been modified in a different way from the native cluster, an effectively combinatorial library of polyketides can be produced as a result of the multiple variations in these activities. As will be further described below, the metes and bounds of this embodiment of the invention can be described on the polyketide, protein, and the encoding nucleotide sequence levels.

As described above, a modular PKS "derived from" the oleandolide or other naturally occurring PKS includes a modular PKS (or its corresponding encoding gene(s)) that retains the scaffolding of the utilized portion of the naturally occurring gene. Not all modules need be included in the constructs; the constructs can include a loading module and six, fewer than six, or more than six extender modules. On the constant scaffold, at least one enzymatic activity is mutated, deleted, replaced, or inserted so as to alter the activity of the resulting PKS relative to the original PKS. Alteration results when these activities are deleted or are replaced by a different version of the activity, or simply mutated in such a way that a polyketide other than the natural product results from these collective activities. This occurs because there has been a resulting alteration of the starter unit and/or extender unit, stereochemistry, chain length or cyclization, and/or reductive or dehydration cycle outcome at a corresponding position in the product polyketide. Where a deleted activity is replaced, the origin of the replacement activity may come from a corresponding activity in a different naturally occurring PKS or from a different region of the oleandolide PKS. Any or all of the oleandolide PKS genes may be included in the derivative or portions of any of these may be included, but the scaffolding of the PKS protein is retained in whatever derivative is constructed. The derivative preferably contains a thioesterase activity from the oleandolide or another PKS.

Thus, a PKS derived from the oleandolide PKS includes a PKS that contains the scaffolding of all or a portion of the oleandolide PKS. The derived PKS also

contains at least two extender modules that are functional, preferably three extender modules, and more preferably four or more extender modules, and most preferably six extender modules. The derived PKS also contains mutations, deletions, insertions, or replacements of one or more of the activities of the functional modules of the oleandolide PKS so that the nature of the resulting polyketide is altered at both the protein and DNA sequence levels. Particular preferred embodiments include those wherein a KS, AT, or ACP domain has been deleted or replaced by a version of the activity from a different PKS or from another location within the same PKS. Also preferred are derivatives where at least one non-condensation cycle enzymatic activity (KR, DH, or ER) has been deleted or added or wherein any of these activities has been mutated so as to change the structure of the polyketide synthesized by the PKS.

Conversely, also included within the definition of a PKS derived from the oleandolide PKS are functional non-oleandolide PKS modules or their encoding genes wherein at least one portion, or two or more portions, of the oleandolide PKS activities have been inserted. Exemplary is the use of the oleandolide AT for extender module 2, which accepts a methylmalonyl CoA extender unit rather than malonyl CoA, to replace a malonyl specific AT in another PKS. Other examples include insertion of portions of non-condensation cycle enzymatic activities or other regions of oleandolide synthase activity into a heterologous PKS at both the DNA and protein levels.

Thus, there are at least five degrees of freedom for constructing a hybrid PKS in terms of the polyketide that will be produced. First, the polyketide chain length is determined by the number of modules in the PKS, and the present invention includes hybrid PKSs that contain a loading module and 6, as well as fewer or more than 6, extender modules. Second, the nature of the carbon skeleton of the PKS is determined by the specificities of the acyl transferases that determine the nature of the extender units at each position, e.g., malonyl, methylmalonyl, ethylmalonyl, or other substituted malonyl. Third, the loading module specificity also has an effect on the resulting carbon skeleton of the polyketide. The loading module may use a different starter unit, such as priopionyl, butyryl, and the like. As noted above and in the examples below, another method for varying loading module specificity involves inactivating the KS activity in extender module 1 (KS1) and providing alternative substrates, called diketides, that are chemically synthesized analogs of extender

module 1 diketide products, for extender module 2. This approach was illustrated in PCT publication Nos. 97/02358 and 99/03986, incorporated herein by reference, wherein the KS1 activity was inactivated through mutation. Fourth, the oxidation state at various positions of the polyketide will be determined by the dehydratase and reductase portions of the modules. This will determine the presence and location of ketone and alcohol moieties and C-C double bonds or C-C single bonds in the polyketide. Finally, the stereochemistry of the resulting polyketide is a function of three aspects of the synthase. The first aspect is related to the AT/KS specificity associated with substituted malonyls as extender units, which affects stereochemistry only when the reductive cycle is missing or when it contains only a ketoreductase, as the dehydratase would abolish chirality. Second, the specificity of the ketoreductase may determine the chirality of any beta-OH. Finally, the enoylreductase specificity for substituted malonyls as extender units may influence the stereochemistry when there is a complete KR/DH/ER available.

Thus, the modular PKS systems generally and the oleandolide PKS system particularly permit a wide range of polyketides to be synthesized. As compared to the aromatic PKS systems, the modular PKS systems accept a wider range of starter units, including aliphatic monomers (acetyl, propionyl, butyryl, isovaleryl, etc.), aromatics (aminohydroxybenzoyl), alicyclics (cyclohexanoyl), and heterocyclics (thiazolyl). Certain modular PKSs have relaxed specificity for their starter units (Kao *et al.*, 1994, *Science*, *supra*). Modular PKSs also exhibit considerable variety with regard to the choice of extender units in each condensation cycle. The degree of beta-ketoreduction following a condensation reaction has also been shown to be altered by genetic manipulation (Donadio *et al.*, 1991, *Science*, *supra*; Donadio *et al.*, 1993, *Proc. Natl. Acad. Sci. USA 90*: 7119-7123). Likewise, the size of the polyketide product can be varied by designing mutants with the appropriate number of modules (Kao *et al.*, 1994, *J. Am. Chem. Soc. 116*:11612-11613). Lastly, modular PKS enzymes are particularly well known for generating an impressive range of asymmetric centers in their products in a highly controlled manner. The polyketides, antibiotics, and other compounds produced by the methods of the invention are typically single stereoisomeric forms. Although the compounds of the invention can occur as mixtures of stereoisomers, it may be beneficial in some instances to generate individual stereoisomers. Thus, the combinatorial potential within modular PKS pathways based

on any naturally occurring modular, such as the oleandolide, PKS scaffold is virtually unlimited.

While hybrid PKSs are most often produced by "mixing and matching" portions of PKS coding sequences, mutations in DNA encoding a PKS can also be used to introduce, alter, or delete an activity in the encoded polypeptide. Mutations can be made to the native sequences using conventional techniques. The substrates for mutation can be an entire cluster of genes or only one or two of them; the substrate for mutation may also be portions of one or more of these genes. Techniques for mutation include preparing synthetic oligonucleotides including the mutations and inserting the mutated sequence into the gene encoding a PKS subunit using restriction endonuclease digestion. See, *e.g.*, Kunkel, 1985, *Proc. Natl. Acad. Sci. USA 82*: 448; Geisselsoder *et al.*, 1987, *BioTechniques 5*:786. Alternatively, the mutations can be effected using a mismatched primer (generally 10-20 nucleotides in length) that hybridizes to the native nucleotide sequence, at a temperature below the melting temperature of the mismatched duplex. The primer can be made specific by keeping primer length and base composition within relatively narrow limits and by keeping the mutant base centrally located. See Zoller and Smith, 1983, *Methods Enzymol. 100*:468. Primer extension is effected using DNA polymerase, the product cloned, and clones containing the mutated DNA, derived by segregation of the primer extended strand, selected. Identification can be accomplished using the mutant primer as a hybridization probe. The technique is also applicable for generating multiple point mutations. *See, e.g.*, Dalbie-McFarland *et al.*, 1982, *Proc. Natl. Acad. Sci. USA 79*: 6409. PCR mutagenesis can also be used to effect the desired mutations.

Random mutagenesis of selected portions of the nucleotide sequences encoding enzymatic activities can also be accomplished by several different techniques known in the art, e.g., by inserting an oligonucleotide linker randomly into a plasmid, by irradiation with X-rays or ultraviolet light, by incorporating incorrect nucleotides during *in vitro* DNA synthesis, by error-prone PCR mutagenesis, by preparing synthetic mutants, or by damaging plasmid DNA *in vitro* with chemicals. Chemical mutagens include, for example, sodium bisulfite, nitrous acid, nitrosoguanidine, hydroxylamine, agents which damage or remove bases thereby preventing normal base-pairing such as hydrazine or formic acid, analogues of nucleotide precursors such as 5-bromouracil, 2-aminopurine, or acridine intercalating

agents such as proflavine, acriflavine, quinacrine, and the like. Generally, plasmid DNA or DNA fragments are treated with chemical mutagens, transformed into *E. coli* and propagated as a pool or library of mutant plasmids.

In constructing a hybrid PKS of the invention, regions encoding enzymatic activity, i.e., regions encoding corresponding activities from different PKS synthases or from different locations in the same PKS, can be recovered, for example, using PCR techniques with appropriate primers. By "corresponding" activity encoding regions is meant those regions encoding the same general type of activity. For example, a KR activity encoded at one location of a gene cluster "corresponds" to a KR encoding activity in another location in the gene cluster or in a different gene cluster. Similarly, a complete reductase cycle could be considered corresponding. For example, KR/DH/ER can correspond to a KR alone.

If replacement of a particular target region in a host PKS is to be made, this replacement can be conducted *in vitro* using suitable restriction enzymes. The replacement can also be effected *in vivo* using recombinant techniques involving homologous sequences framing the replacement gene in a donor plasmid and a receptor region in a recipient plasmid. Such systems, advantageously involving plasmids of differing temperature sensitivities are described, for example, in PCT publication No. WO 96/40968, incorporated herein by reference. The vectors used to perform the various operations to replace the enzymatic activity in the host PKS genes or to support mutations in these regions of the host PKS genes can be chosen to contain control sequences operably linked to the resulting coding sequences in a manner such that expression of the coding sequences can be effected in an appropriate host.

However, simple cloning vectors may be used as well. If the cloning vectors employed to obtain PKS genes encoding derived PKS lack control sequences for expression operably linked to the encoding nucleotide sequences, the nucleotide sequences are inserted into appropriate expression vectors. This need not be done individually, but a pool of isolated encoding nucleotide sequences can be inserted into expression vectors, the resulting vectors transformed or transfected into host cells, and the resulting cells plated out into individual colonies. The invention provides a variety of recombinant DNA compounds in which the various coding sequences for the

domains and modules of the oleandolide PKS are flanked by non-naturally occurring restriction enzyme recognition sites.

The various PKS nucleotide sequences can be cloned into one or more recombinant vectors as individual cassettes, with separate control elements, or under the control of, e.g., a single promoter. The PKS subunit encoding regions can include flanking restriction sites to allow for the easy deletion and insertion of other PKS subunit encoding sequences so that hybrid PKSs can be generated. The design of such unique restriction sites is known to those of skill in the art and can be accomplished using the techniques described above, such as site-directed mutagenesis and PCR.

The expression vectors containing nucleotide sequences encoding a variety of PKS enzymes for the production of different polyketides are then transformed into the appropriate host cells to construct the library. In one straightforward approach, a mixture of such vectors is transformed into the selected host cells and the resulting cells plated into individual colonies and selected to identify successful transformants. Each individual colony has the ability to produce a particular PKS synthase and ultimately a particular polyketide. Typically, there will be duplications in some, most, or all of the colonies; the subset of the transformed colonies that contains a different PKS in each member colony can be considered the library. Alternatively, the expression vectors can be used individually to transform hosts, which transformed hosts are then assembled into a library. A variety of strategies are available to obtain a multiplicity of colonies each containing a PKS gene cluster derived from the naturally occurring host gene cluster so that each colony in the library produces a different PKS and ultimately a different polyketide. The number of different polyketides that are produced by the library is typically at least four, more typically at least ten, and preferably at least 20, and more preferably at least 50, reflecting similar numbers of different altered PKS gene clusters and PKS gene products. The number of members in the library is arbitrarily chosen; however, the degrees of freedom outlined above with respect to the variation of starter, extender units, stereochemistry, oxidation state, and chain length enables the production of quite large libraries.

Methods for introducing the recombinant vectors of the invention into suitable hosts are known to those of skill in the art and typically include the use of $CaCl_2$ or agents such as other divalent cations, lipofection, DMSO, PEG, protoplast transformation, infection, transfection, and electroporation. The polyketide producing

colonies can be identified and isolated using known techniques and the produced polyketides further characterized. The polyketides produced by these colonies can be used collectively in a panel to represent a library or may be assessed individually for activity.

The libraries of the invention can thus be considered at four levels: (1) a multiplicity of colonies each with a different PKS encoding sequence; (2) the proteins produced from the coding sequences; (3) the polyketides produced from the proteins assembled into a functional PKS; and (4) antibiotics or compounds with other desired activities derived from the polyketides. Combination libraries can also be constructed wherein members of a library derived, for example, from the oleandolide PKS can be considered as a part of the same library as those derived from, for example, the rapamycin PKS or DEBS.

Colonies in the library are induced to produce the relevant synthases and thus to produce the relevant polyketides to obtain a library of polyketides. Polyketides that are secreted into the media or have been otherwise isolated can be screened for binding to desired targets, such as receptors, signaling proteins, and the like. The supernatants *per se* can be used for screening, or partial or complete purification of the polyketides can first be effected. Typically, such screening methods involve detecting the binding of each member of the library to receptor or other target ligand. Binding can be detected either directly or through a competition assay. Means to screen such libraries for binding are well known in the art. Alternatively, individual polyketide members of the library can be tested against a desired target. In this event, screens wherein the biological response of the target is measured can more readily be included. Antibiotic activity can be verified using typical screening assays such as those set forth in Lehrer *et al.*, 1991, *J. Immunol. Meth. 137*:167-173, incorporated herein by reference, and in Example 7, below.

The invention provides methods for the preparation of a large number of polyketides. These polyketides are useful intermediates in formation of compounds with antibiotic or other activity through hydroxylation and glycosylation reactions as described above. In general, the polyketide products of the PKS must be further modified, typically by hydroxylation and glycosylation, to exhibit antibiotic activity. Hydroxylation results in the novel polyketides of the invention that contain hydroxyl groups at C-6, which can be accomplished using the hydroxylase encoded by the *eryF*

gene, and/or C-12, which can be accomplished using the hydroxylase encoded by the *picK* or *eryK* gene. Also, the present invention provides the *oleP* gene in recombinant form, which can be used to express the *oleP* gene product in any host cell. A host cell, such as a *Streptomyces* host cell or a *Saccharopolyspora erythraea* host cell modified to express the *oleP* gene thus can be used to produce polyketides comprising the C-8-C-8a epoxide present in oleandomycin. Thus the invention provides such modified polyketides. The presence of hydroxyl groups at these positions can enhance the antibiotic activity of the resulting compound relative to its unhydroxylated counterpart.

Methods for glycosylating the polyketides are generally known in the art; the glycosylation may be effected intracellularly by providing the appropriate glycosylation enzymes or may be effected *in vitro* using chemical synthetic means as described herein and in PCT publication No. WO 98/49315, incorporated herein by reference. Preferably, glycosylation with desosamine is effected in accordance with the methods of the invention in recombinant host cells provided by the invention. In general, the approaches to effecting glycosylation mirror those described above with respect to hydroxylation. The purified enzymes, isolated from native sources or recombinantly produced may be used *in vitro*. Alternatively and as noted, glycosylation may be effected intracellularly using endogenous or recombinantly produced intracellular glycosyl transferases. In addition, synthetic chemical methods may be employed.

The antibiotic modular polyketides may contain any of a number of different sugars, although D-desosamine, or a close analog thereof, is most common. Erythromycin, picromycin, narbomycin, and methymycin contain desosamine. Erythromycin also contains L-cladinose (3-O-methyl mycarose). Tylosin contains mycaminose (4-hydroxy desosamine), mycarose and 6-deoxy-D-allose. 2-acetyl-1-bromodesosamine has been used as a donor to glycosylate polyketides by Masamune *et al.*, 1975, *J. Am. Chem. Soc. 97*: 3512-3513. Other, apparently more stable donors include glycosyl fluorides, thioglycosides, and trichloroacetimidates; see Woodward *et al.*, 1981, *J. Am. Chem. Soc. 103*: 3215; Martin *et al.*, 1997, *J. Am. Chem. Soc. 119*: 3193; Toshima *et al.*, 1995, *J. Am. Chem. Soc. 117*: 3717; Matsumoto *et al.*, 1988, *Tetrahedron Lett. 29*: 3575. Glycosylation can also be effected using the polyketide aglycones as starting materials and using *Saccharopolyspora erythraea*,*Streptomyces*

venezuelae or other host cells to make the conversion, preferably using mutants unable to synthesize macrolides, as discussed in the preceding Section.

Thus, a wide variety of polyketides can be produced by the hybrid PKS enzymes of the invention. These polyketides are useful as antibiotics and as intermediates in the synthesis of other useful compounds, as described in the following section.

Section IV: Compounds

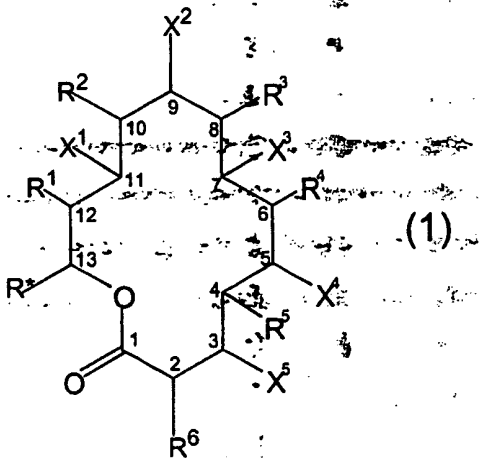
The methods and recombinant DNA compounds of the invention are useful in the production of polyketides. In one important aspect, the invention provides methods for making antibiotic compounds related in structure to oleandomycin and erythromycin, both potent antibiotic compounds. The invention also provides novel ketolide compounds, polyketide compounds with potent antibiotic activity of significant interest due to activity against antibiotic resistant strains of bacteria. See Griesgraber *et al.*, 1996, *J. Antibiot. 49*: 465-477, incorporated herein by reference. Most if not all of the ketolides prepared to date are synthesized using erythromycin A, a derivative of 6-dEB, as an intermediate. While the invention provides hybrid PKSs that produce a polyketide different in structure from 6-dEB, the invention also provides methods for making intermediates useful in preparing traditional, 6-dEB- and erythromycin-derived ketolide compounds.

Because 6-dEB in part differs from oleandolide in that it comprises a 13-ethyl instead of a 13-methyl group, the novel hybrid PKS genes of the invention based on the oleandolide PKS provide many novel ketolides that differ from the known ketolides only in that they have a 13-methyl instead of 13-ethyl group. Thus, the invention provides the 13-methyl analogues of the ketolides and intermediates and precursor compounds described in, for example, Griesgraber *et al., supra*; Agouridas *et al.*, 1998, *J. Med. Chem. 41*: 4080-4100, U.S. Patent Nos. 5,770,579; 5,760,233; 5,750,510; 5,747,467; 5,747,466; 5,656,607; 5,635,485; 5,614,614; 5,556,118; 5,543,400; 5,527,780; 5,444,051; 5,439,890; 5,439,889; and PCT publication Nos. WO 98/09978 and 98/28316, each of which is incorporated herein by reference.

As noted above, the hybrid PKS genes of the invention can be expressed in a host cell that contains the desosamine biosynthetic genes and desosaminyl transferase gene as well as the required hydroxylase gene(s), which may be either *picK* (for the

C-12 position) or *eryK* (for the C-12 position) and/or *eryF* (for the C-6 position). The resulting compounds have antibiotic activity but can be further modified, as described in the patent publications referenced above, to yield a desired compound with improved or otherwise desired properties. Alternatively, the aglycone compounds can be produced in the recombinant host cell, and the desired glycosylation and hydroxylation steps carried out *in vitro* or *in vivo*, in the latter case by supplying the converting cell with the aglycone.

The compounds of the invention are thus optionally glycosylated forms of the polyketide set forth in formula (1) below which are hydroxylated at either the C-6 or the C-12 or both. The compounds of formula (1) can be prepared using the loading and the six extender modules of a modular PKS, modified or prepared in hybrid form as herein described. These polyketides have the formula:



(1)

including the glycosylated and isolated stereoisomeric forms thereof;

wherein R* is a straight chain, branched or cyclic, saturated or unsaturated substituted or unsubstituted hydrocarbyl of 1-15C;

each of R^1-R^6 is independently H or alkyl (1-4C) wherein any alkyl at R^1 may optionally be substituted;

each of X^1-X^5 is independently two H, H and OH, or =O; or

each of X^1-X^5 is independently H and the compound of formula (2) contains a double-bond in the ring adjacent to the position of said X at 2-3, 4-5, 6-7, 8-9 and/or 10-11;

with the proviso that:

at least two of R^1-R^6 are alkyl (1-4C).

Preferred compounds comprising formula 2 are those wherein at least three of R^1-R^5 are alkyl (1-4C), preferably methyl or ethyl; more preferably wherein at least four of R^1-R^5 are alkyl (1-4C), preferably methyl or ethyl. Also preferred are those wherein X^2 is two H, =O, or H and OH, and/or X^3 is H, and/or X^1 is OH and/or X^4 is OH and/or X^5 is OH. Also preferred are compounds with variable R* when R^1-R^5 is methyl, X^2 is =O, and X^1, X^4 and X^5 are OH. The glycosylated forms of the foregoing are also preferred; glycoside residues can be attached at C-3, C-5, and/or C-6; the epoxidated forms are also included, i.e., and epoxide at C-8-C-8a.

As described above, there are a wide variety of diverse organisms that can modify compounds such as those described herein to provide compounds with or that can be readily modified to have useful activities. For example, *Saccharopolyspora erythraea* can convert oleandolide and 6-dEB to a variety of useful compounds. The compounds provided by the present invention can be provided to cultures of *Saccharopolyspora erythraea* and converted to the corresponding derivatives of erythromycins A, B, C, and D in accordance with the procedure provided in Example 6, below. To ensure that only the desired compound is produced, one can use an *S. erythraea eryA* mutant that is unable to produce 6-dEB but can still carry out the desired conversions (Weber *et al.*, 1985, *J. Bacteriol. 164*(1): 425-433). Also, one can employ other mutant strains, such as *eryB*, *eryC*, *eryG*, and/or *eryK* mutants, or mutant strains having mutations in multiple genes, to accumulate a preferred compound. The conversion can also be carried out in large fermentors for commercial production. Each of the erythromycins A, B, C, and D has antibiotic activity, although erythromycin A has the highest antibiotic activity. Moreover, each of these compounds can form, under treatment with mild acid, a C-6 to C-9 hemiketal with motilide activity. For formation of hemiketals with motilide activity, erythromycins B, C, and D, are preferred, as the presence of a C-12 hydroxyl allows the formation of an inactive compound that has a hemiketal formed between C-9 and C-12.

Thus, the present invention provides the compounds produced by hydroxylation and glycosylation of the compounds of the invention by action of the enzymes endogenous to *Saccharopolyspora erythraea* and mutant strains of *S. erythraea*. Such compounds are useful as antibiotics or as motilides directly or after chemical modification. For use as antibiotics, the compounds of the invention can be used directly without further chemical modification. Erythromycins A, B, C, and D all

have antibiotic activity, and the corresponding compounds of the invention that result from the compounds being modified by *Saccharopolyspora erythraea* also have antibiotic activity. These compounds can be chemically modified, however, to provide other compounds of the invention with potent antibiotic activity. For example, alkylation of erythromycin at the C-6 hydroxyl can be used to produce potent antibiotics (clarithromycin is C-6-O-methyl), and other useful modifications are described in, for example, Griesgraber *et al.*, 1996, *J. Antibiot. 49*: 465-477, Agouridas *et al.*, 1998, *J. Med. Chem. 41*: 4080-4100, U.S. Patent Nos. 5,770,579; 5,760,233; 5,750,510; 5,747,467; 5,747,466; 5,656,607; 5,635,485; 5,614,614; 5,556,118; 5,543,400; 5,527,780; 5,444,051; 5,439,890; and 5,439,889; and PCT publication Nos. WO 98/09978 and 98/28316, each of which is incorporated herein by reference.

For use as motilides, the compounds of the invention can be used directly without further chemical modification. Erythromycin and certain erythromycin analogs are potent agonists of the motilin receptor that can be used clinically as prokinetic agents to induce phase III of migrating motor complexes, to increase esophageal peristalsis and LES pressure in patients with GERD, to accelerate gastric emptying in patients with gastric paresis, and to stimulate gall bladder contractions in patients after gallstone removal and in diabetics with autonomic neuropathy. See Peeters, 1999, Motilide Web Site, http://www.med.kuleuven.ac.be/med/gih/motilid.htm, and Omura *et al.*, 1987, Macrolides with gastrointestinal motor stimulating activity, *J. Med. Chem. 30*: 1941-3). The corresponding compounds of the invention that result from the compounds of the invention being modified by *Saccharopolyspora erythraea* also have motilide activity, particularly after conversion, which can also occur *in vivo*, to the C-6 to C-9 hemiketal by treatment with mild acid. Compounds lacking the C-12 hydroxyl are especially preferred for use as motilin agonists. These compounds can also be further chemically modified, however, to provide other compounds of the invention with potent motilide activity.

Moreover, and also as noted above, there are other useful organisms that can be employed to hydroxylate and/or glycosylate the compounds of the invention. As described above, the organisms can be mutants unable to produce the polyketide normally produced in that organism, the fermentation can be carried out on plates or in large fermentors, and the compounds produced can be chemically altered after

fermentation. In addition to *Saccharopolyspora erythraea*, *Streptomyces venezuelae*, *S. narbonensis*, *S. antibioticus*, *Micromonospora megalomicea*, *S. fradiae*, and *S. thermotolerans* can also be used. In addition to antibiotic activity, compounds of the invention produced by treatment with *M. megalomicea* enzymes can have antiparasitic activity as well. Thus, the present invention provides the compounds produced by hydroxylation and glycosylation by action of the enzymes endogenous to *S. erythraea*, *S. venezuelae*, *S. narbonensis*, *S. antibioticus*, *M. megalomicea*, *S. fradiae*, and *S. thermotolerans*.

The present invention also provides methods and genetic constructs for producing the glycosylated and/or hydroxylated compounds of the invention directly in the host cell of interest. Thus, the recombinant genes of the invention, which include recombinant *oleAI*, *oleAII*, and *oleAIII* genes with one or more deletions and/or insertions, including replacements of an *oleA* gene fragment with a gene fragment from a heterologous PKS gene, can be included on expression vectors suitable for expression of the encoded gene products in *Saccharopolyspora erythraea*, *Micromonospora megalomicea*, *Streptomyces antibioticus*, *S. venezuelae*, *S. narbonensis*, *S. fradiae*, and *S. thermotolerans*.

Many of the compounds of the invention contain one or more chiral centers, and all of the stereoisomers are included within the scope of the invention, as pure compounds as well as mixtures of stereoisomers. Thus the compounds of the invention may be supplied as a mixture of stereoisomers in any proportion.

The compounds of the invention can be produced by growing and fermenting the host cells of the invention under conditions known in the art for the production of other polyketides. The compounds of the invention can be isolated from the fermentation broths of these cultured cells and purified by standard procedures. The compounds can be readily formulated to provide the pharmaceutical compositions of the invention. The pharmaceutical compositions of the invention can be used in the form of a pharmaceutical preparation, for example, in solid, semisolid, or liquid form. This preparation will contain one or more of the compounds of the invention as an active ingredient in admixture with an organic or inorganic carrier or excipient suitable for external, enteral, or parenteral application. The active ingredient may be compounded, for example, with the usual non-toxic, pharmaceutically acceptable

carriers for tablets, pellets, capsules, suppositories, solutions, emulsions, suspensions, and any other form suitable for use.

The carriers which can be used include water, glucose, lactose, gum acacia, gelatin, mannitol, starch paste, magnesium trisilicate, talc, corn starch, keratin, colloidal silica, potato starch, urea, and other carriers suitable for use in manufacturing preparations, in solid, semi-solid, or liquified form. In addition, auxiliary stabilizing, thickening, and coloring agents and perfumes may be used. For example, the compounds of the invention may be utilized with hydroxypropyl methylcellulose essentially as described in U.S. Patent No. 4,916,138, incorporated herein by reference, or with a surfactant essentially as described in EPO patent publication No. 428,169, incorporated herein by reference.

Oral dosage forms may be prepared essentially as described by Hondo *et al.*, 1987, *Transplantation Proceedings XIX*, Supp. 6: 17-22, incorporated herein by reference. Dosage forms for external application may be prepared essentially as described in EPO patent publication No. 423,714, incorporated herein by reference. The active compound is included in the pharmaceutical composition in an amount sufficient to produce the desired effect upon the disease process or condition.

For the treatment of conditions and diseases caused by infection, a compound of the invention may be administered orally, topically, parenterally, by inhalation spray, or rectally in dosage unit formulations containing conventional non-toxic pharmaceutically acceptable carriers, adjuvant, and vehicles. The term parenteral, as used herein, includes subcutaneous injections, and intravenous, intramuscular, and intrasternal injection or infusion techniques.

Dosage levels of the compounds of the invention are of the order from about 0.01 mg to about 50 mg per kilogram of body weight per day, preferably from about 0.1 mg to about 10 mg per kilogram of body weight per day. The dosage levels are useful in the treatment of the above-indicated conditions (from about 0.7 mg to about 3.5 mg per patient per day, assuming a 70 kg patient). In addition, the compounds of the invention may be administered on an intermittent basis, i.e., at semi-weekly, weekly, semi-monthly, or monthly intervals.

The amount of active ingredient that may be combined with the carrier materials to produce a single dosage form will vary depending upon the host treated and the particular mode of administration. For example, a formulation intended for

oral administration to humans may contain from 0.5 mg to 5 gm of active agent compounded with an appropriate and convenient amount of carrier material, which may vary from about 5 percent to about 95 percent of the total composition. Dosage unit forms will generally contain from about 0.5 mg to about 500 mg of active ingredient. For external administration, the compounds of the invention may be formulated within the range of, for example, 0.00001% to 60% by weight, preferably from 0.001% to 10% by weight, and most preferably from about 0.005% to 0.8% by weight.

It will be understood, however, that the specific dose level for any particular patient will depend on a variety of factors. These factors include the activity of the specific compound employed; the age, body weight, general health, sex, and diet of the subject; the time and route of administration and the rate of excretion of the drug; whether a drug combination is employed in the treatment; and the severity of the particular disease or condition for which therapy is sought.

The compounds of the invention can be used as single therapeutic agents or in combination with other therapeutic agents. Drugs that can be usefully combined with compounds of the invention include one or more antibiotic or motilide agents.

A detailed description of the invention having been provided above, the following examples are given for the purpose of illustrating the invention and shall not be construed as being a limitation on the scope of the invention or claims.

Example 1

General Methodology

Bacterial strains, plasmids, and culture conditions. *Streptomyces coelicolor* CH999 described in WO 95/08548, published 30 March 1995, or *S. lividans* K4-114 or K4-155, described in Ziermann and Betlach, Jan. 99, Recombinant Polyketide Synthesis in *Streptomyces*: Engineering of Improved Host Strains, *BioTechniques 26*:106-110, incorporated herein by reference, was used as an expression host. DNA manipulations were performed in *Escherichia coli* XL1-Blue, available from Stratagene. *E. coli* MC1061 is also suitable for use as a host for plasmid manipulation. Plasmids were passaged through *E. coli* ET12567 (*dam dcm hsdS* Cm^r) (MacNeil, 1988, *J. Bacteriol. 170*: 5607, incorporated herein by reference) to generate unmethylated DNA prior to transformation of *S. coelicolor* or

Saccharopolyspora erythraea. *E. coli* strains were grown under standard conditions. *S. coelicolor* strains were grown on R2YE agar plates (Hopwood *et al., Genetic manipulation of Streptomyces. A laboratory manual.* The John Innes Foundation: Norwich, 1985, incorporated herein by reference).

Many of the expression vectors of the invention illustrated in the examples are derived from plasmid pRM5, described in WO 95/08548, incorporated herein by reference. This plasmid includes a colEI replicon, an appropriately truncated SCP2* *Streptomyces* replicon, two *act*-promoters, the *actI* and *actIII* promoters, to allow for bidirectional cloning, the gene encoding the *actII-ORF4* activator which induces transcription from *act* promoters during the transition from growth phase to stationary phase, and appropriate marker genes. Engineered restriction sites in the plasmid facilitate the combinatorial construction of PKS gene clusters starting from cassettes encoding individual domains of naturally occurring PKSs. When plasmid pRM5 is used for expression of a PKS, all relevant biosynthetic genes can be plasmid-borne and therefore amenable to facile manipulation and mutagenesis in *E. coli*. This plasmid is also suitable for use in *Streptomyces* host cells. *Streptomyces* is genetically and physiologically well characterized and expresses the ancillary activities required for *in vivo* production of most polyketides. Plasmid pRM5 utilizes the *act* promoter for PKS gene expression, so polyketides are produced in a secondary metabolite-like manner, thereby alleviating the toxic effects of synthesizing potentially bioactive compounds *in vivo*.

Manipulation of DNA and organisms. Polymerase chain reaction (PCR) was performed using *Pfu* polymerase (Stratagene; *Taq* polymerase from Perkin Elmer Cetus can also be used) under conditions recommended by the enzyme manufacturer. Standard *in vitro* techniques were used for DNA manipulations (Sambrook *et al. Molecular Cloning: A Laboratory Manual* (Current Edition)). *E. coli* was transformed using standard calcium chloride-based methods; a Bio-Rad *E. coli* pulsing apparatus and protocols provided by Bio-Rad could also be used. *S. coelicolor* was transformed by standard procedures (Hopwood *et al. Genetic manipulation of Streptomyces. A laboratory manual.* The John Innes Foundation: Norwich, 1985), and depending on what selectable marker was employed, transformants were selected using 1 mL of a 1.5 mg/mL thiostrepton overlay, 1 mL of a 2 mg/mL apramycin overlay, or both.

Example 2

Cloning of the Oleandomycin Biosynthetic Gene Cluster from *Streptomyces antibioticus*

Genomic DNA (100 μg) was isolated from an oleandomycin producing strain of *Streptomyces antibioticus* (ATCC 11891) using standard procedures. The genomic DNA was partially digested with restriction enzyme *Sau*3A1 to generate fragments ~40 kbp in length, which were cloned into the commercially available Supercos™ cosmid vector that had been digested with restriction enzymes *Xba*I and *Bam*HI to produce a genomic library. SuperCosI™ (Stratagene) DNA cosmid arms were prepared as directed by the manufacturer. A cosmid library was prepared by ligating 2.5 μg of the digested genomic DNA with 1.5 μg of cosmid arms in a 20 μL reaction. One microliter of the ligation mixture was propagated in *E. coli* XL1-Blue MR (Stratagene) using a GigapackIII XL packaging extract kit (Stratagene).

This library was then probed with a radioactively-labeled probe generated by PCR from *Streptomyces antibioticus* DNA using primers complementary to known sequences of KS domains hypothesized to originate from extender modules 5 and 6 of the oleandolide PKS. This probing identified about 30 different colonies, which were pooled, replated, and probed again, resulting in the identification of 9 cosmids. These latter cosmids were isolated and transformed into the commercially available *E. coli* strain XL-1 Blue. Plasmid DNA was isolated and analyzed by restriction enzyme digestion, which revealed that the entire PKS gene cluster was contained in overlapping segments on two of the cosmids identified. DNA sequence analysis using the T3 primer showed that the desired DNA had been isolated.

Further analysis of these cosmids and subclones prepared from the cosmids facilitated the identification of the location of various oleandolide PKS ORFs, modules in those ORFs, and coding sequences for oleandomycin modification enzymes. The location of these genes and modules is shown on Figure 1. Figure 1 shows that the complete oleandolide PKS gene cluster is contained within the insert DNA of cosmids pKOS055-1 (insert size of ~43 kb) and pKOS055-5 (insert size of ~47 kb). Each of these cosmids has been deposited with the American Type Culture Collection in accordance with the terms of the Budapest Treaty (cosmid pKOS055-1 is available under accession no. ATCC 203798; cosmid pKOS055-5 is available under accession no. ATCC 203799). Various additional reagents of the invention can

therefore be isolated from these cosmids. DNA sequence analysis was also performed on the various subclones of the invention, as described above.

Example 3

Expression of an Oleandolide/DEBS Hybrid PKS in *Saccharopolyspora erythraea*

This Example describes the construction of an expression vector, plasmid pKOS039-110, that can integrate into the chromosome of *Saccharopolyspora erythraea* due to the phage phiC31 attachment and integration functions present on the plasmid and drive expression of the *oleAI* gene product under the control of the *ermE** promoter. A restriction site and function map of plasmid pKOS039-110 is shown in Figure 3 of the accompanying drawings. The expression of the *oleAI* gene product in a host cell that naturally produces the *eryA* gene products results in the formation of a functional hybrid PKS of the present invention composed of the *oleAI*, *eryAII*, and *eryAIII* gene products and the concomitant production of 13-methyl erythromycins. While the specific plasmids and vectors utilized in the construction are described herein, those of skill in the art will recognize that equivalent expression vectors of the invention can be readily constructed from publicly available materials and the *oleA* gene containing cosmids of the present invention deposited with the ATCC.

Plasmid pKOS039-98 is a cloning vector that contains convenient restriction sites that was constructed by inserting a polylinker oligonucleotide, containing a restriction enzyme recognition site for *Pac*I, a Shine-Dalgarno sequence, and restriction enzyme recognition sites for *Nde*I, *Bgl*II, and *Hin*dIII, into a pUC19 derivative, called pKOS24-47. Plasmid pKOS039-98 (see PCT patent application No. WO US99/11814, incorporated herein by reference) was digested with restriction enzymes *Pac*I and *Eco*RI and ligated to a polylinker composed of the oligonucleotides N39-51 and N39-52 having the following sequence:

N39-51: 5'-TAAGGAGGACCATATGCATCGCTCGAGTCTAGACCTAGG-3'

N39-52: 5'-AATTCCTAGGTCTAGACTCGAGCGATGCATATGGTCCTCC-TTAAT-3', which thus includes the following restriction enzyme recognition sites in the order shown: *Nde*I-*Nsi*I-*Xho*I-*Xba*I-*Eco*RI, to yield plasmid pKOS039-105.

Plasmid pKOS039-105 was digested with restriction enzymes *Nsi*I and *Eco*RI, and the resulting large fragment ligated to the 15.2 kb *Nsi*I-*Eco*RI restriction fragment of cosmid pKOS055-5 containing the *oleAI* gene to yield plasmid pKOS039-116.

Plasmid pKOS039-116 was digested with restriction enzymes *Nde*I and *Eco*RI, and the resulting 15.2 kb fragment containing the *oleAI* gene was isolated and ligated to the 6 kb *Nde*I-*Eco*RI restriction fragment of plasmid pKOS039-134B to yield plasmid pKOS039-110 (Figure 3).

Plasmid pKOS039-134B is a derivative of pKOS039-104 described in PCT patent application No. WO US99/11814, *supra*, prepared by digesting the latter with restriction enzyme *Bgl*II and ligating the ~10.5 kb fragment to get pKOS39-104B. Plasmid pKOS39-104B was digested with restriction enzyme *Pac*I and partially digested with restriction enzyme *Xba*I. The ~7.4 kb fragment was ligated with PCR61A+62 fragment treated with restriction enzymes *Pac*I and *Avr*II. The PCR61A+62 fragment was generated using the PCR primers:
N39-61A, 5'-TTCCTAGGCTAGCCCGACCCGAGCACGCGCCGGCA-3'; and
N39-62, 5'-CCTTAATTAAGGATCCTACCAACCGGCACGATTGTGCC-3',
and the template was pWHM1104 (Tang *et al.*, 1996, *Molecular Microbiology* 22(5): 801-813).

Plasmid pKOS039-110 DNA was passed through *E. coli* ET cells to obtain non-methylated DNA, which was then used to transform *Saccharopolyspora erythraea* cells, which contain a mutation in the *eryAI* coding sequence for the KS domain of module 1 of DEBS that renders the PKS non-functional. The resulting transformants produced detectable amounts of 14-desmethyl erythromycins.

Example 4

Heterologous Expression of an Oleandolide PKS in *Streptomyces lividans*

This Example describes the construction of an expression vector, plasmid pKOS039-130, that has an SCP2* origin of replication and so can replicate in *Streptomyces* host cells and drive expression of the *oleAI*, *oleAII*, and *oleAIII* gene products under the control of the *actI* promoter and *actII-ORF4* activator. A restriction site and function map of plasmid pKOS039-130 is shown in Figure 4 of the accompanying drawings. The expression of the *oleA* gene products in this host cell results in the formation of a functional oleandolide PKS composed of the *oleAI*, *oleAII*, and *oleAIII* gene products and the concomitant production of 8,8a-deoxyoleandolide. While the specific plasmids and vectors utilized in the construction are described herein, those of skill in the art will recognize that equivalent expression

vectors of the invention can be readily constructed from publicly available materials and the *oleA* gene containing cosmids of the present invention deposited with the ATCC.

The 7.2 kb *Nsi*I-*Xho*I restriction fragment of cosmid pKOS055-5 was cloned into pKOS39-105 to give plasmid pKOS039-106. The 8.0 kb *Xho*I-*Pst*I restriction fragment of cosmid pKOS055-5 was cloned into commercially available plasmid pLitmus28 to yield plasmid pKOS039-107. The 14 kb *Eco*RI-*Eco*RV and 5.4 kb *Eco*RV-*Pst*I restriction fragments of cosmid pKOS055-1 were ligated with pLitmus28 digested with *Eco*RI and *Pst*I to yield plasmid pKOS039-115. The 19.5 kb *Spe*I-*Xba*I restriction fragment from plasmid pKOS039-115 was inserted into pKOS039-73, a derivative of plasmid pRM5, to yield plasmid pKOS039-129. The 15.2 kb *Pac*I-*Eco*RI restriction fragment of plasmid pKOS039-110 was inserted into pKOS039-129 by replacing the 22 kb *Pac*I - *Eco*RI restriction fragment to yield plasmid pKOS038-174. The 19 kb *Eco*RI restriction fragment from plasmid pKOS039-129 was then inserted into pKOS038-174 to yield plasmid pKOS039-130 (Figure 4), which was used to transform *Streptomyces lividans* K4-114 (K4-155 could also be used). The resulting transformants produced 8,8a-deoxyoleandolide.

As noted above, the invention provides a recombinant *oleAI* gene in which the coding sequence for the KS domain of module 1 has been mutated to change the active site cysteine to another amino acid (the KS1° mutation). Recombinant PKS enzymes comprising this gene product do not produce a polyketide unless provided with diketide (or triketide) compounds that can bind to the KS2 or KS3 domain, where they are then processed to form a polyketide comprising the diketide (or triketide). This recombinant *oleAI* gene can be used together with the *oleAII* and *oleAII* genes to make a recombinant oleandolide PKS or can be used with modified forms of those genes or other naturally occurring or recombinant PKS genes to make a hybrid PKS.

To make the KS1° mutation in *oleAI*, the following primers were prepared:

N39-47, 5'-GCGAATTCCCGGGTGGCGTGACCTCT;

N39-48, 5'-GAGCTAGCCGCCGTGTCCACCGTGACC;

N39-49, 5'-CGGCTAGCTCGTCGCTGGTGGCACTGCAC; and

N39-50, 5'-CGAAGCTTGACCAGGAAAGACGAACACC.

These primers were used to amplify template DNA prepared from pKOS039-106. The amplification product of primers N39-47 and N39-48 was digested with restriction enzymes *Eco*RI and *Nhe*I, and the amplification product of primers N39-49 and N39-50 was digested with restriction enzymes *Nhe*I and *Hin*dIII, and the resulting restriction fragments were ligated to *Eco*RI and *Hin*dIII-digested plasmid pLitmus28 to yield plasmid pKOS038-179. The 1.5 kb *Bsr*GI-*Bbv*CI restriction fragment of plasmid pKOS038-179 was inserted into plasmid pKOS039-106 to yield pKOS098-2. The 7 kb *Nsi*I - *Xho*I restriction fragment of plasmid pKOS098-2 and the 8 kb *Xho*I - *Eco*RI restriction fragments of plasmid pKOS039-107 are then used to replace the 15.2 kb *Nsi*I - *Eco*RI restriction fragment of plasmid pKOS039-110 to yield the desired expression vector, pKOS039-110-KS1°, which comprises the *oleAI* KS1° gene under the control of the *ermE** promoter.

To provide an expression vector of the invention that encodes the complete oleandolide PKS with the recombinant *oleAI* KS1° gene product, the *oleAI* KS1° gene can be isolated as a *Pac*I - *Eco*RI restriction fragment from plasmid pKOS039-110-KS1°, which is then used to construct an expression vector analogous to the expression vector plasmid pKOS039-130 in the same manner in which the latter vector was constructed. The resulting expression vector can be used in *Streptomyces lividans*, *S. coelicolor*, and other compatible host cells to make polyketides by diketide feeding as described in PCT patent publication No. WO 99/03986, incorporated herein by reference.

Example 5

Expression of an Oleandomycin/Picromycin Hybrid PKS

This Example describes the construction of an expression vector, plasmid pKOS039-133, that can integrate into the chromosome of *Streptomyces* due to the phage phiC31 attachment and integration functions present on the plasmid and drive expression of the *oleAIII* gene product under the control of the *actI* promoter and *actII-ORF4* activator. A restriction site and function map of plasmid pKOS039-133 is shown in Figure 5 of the accompanying drawings. This plasmid was introduced into *S. lividans* host cells together with a plasmid, pKOS039-83, that drives expression of the narbonolide PKS genes *picAI* and *picAII* (see PCT patent application No. WO US99/11814, *supra*). The expression of the *oleAIII* and *picAI* and *picAII* gene

products in a host cell results in the formation of a functional hybrid PKS of the present invention composed of the *oleAIII*, *picAI*, and *picAII* gene products and the concomitant production of 3-hydroxy-narbonolide. While the specific plasmids and vectors utilized in the construction are described herein, those of skill in the art will recognize that equivalent expression vectors of the invention can be readily constructed from publicly available materials and the *oleA* gene containing cosmids of the present invention deposited with the ATCC.

Two oligonucleotides were prepared for the insertion of the *oleAIII* gene into pSET152 derivative plasmid pKOS039-42:

N39-59, 5'-AATTCATATGGCTGAGGCGGAGAAGCTGCGCGAATACC-TGTGG; and

N39-60, 5'-CGCGCCACAGGTATTCGCGCAGCTTCTCCGCCTCAGCCATATG.

Plasmid pKOS039-115 was digested with restriction enzymes *Eco*RI and *Asc*I to give the ~13.8 kb restriction fragment, which was inserted with the linker N39-59/N39-60 to yield plasmid pKOS039-132. Plasmid pKOS039-132 was digested with restriction enzymes *Nde*I and *Xba*I to give the ~10.8 kb restriction fragment, which was ligated to the ~9 kb *Nde*I-*Spe*I restriction fragment of plasmid pKOS039-42 to yield plasmid pKOS039-133 (Figure 5). Plasmid pKOS039-133 and pKOS039-83 were co-transformed into *Streptomyces lividans* K4-114 (K4-155 can also be used; see Ziermann and Betlach, 1999, *Biotechniques 26*, 106-110, and U.S. patent application Serial No. 09/181,833, filed 28 Oct. 1998, each of which is incorporated herein by reference). Protoplasts were transformed using standard procedures and transformants selected using overlays containing antibiotics. The strains were grown in liquid R5 medium (with 20 μg/mL thiostrepton, see Hopwood *et al.*, *Genetic Manipulation of Streptomyces: A Laboratory Manual*; John Innes Foundation: Norwich, UK, 1985, incorporated herein by reference) for growth/seed and production cultures at 30°C. Analysis of extracts by LC/MS established the identity of the polyketide as the expected compound, 3-hydroxynarbonolide.

Example 6

Conversion of Erythronolides to Erythromycins

A sample of an oleandolide (~50 to 100 mg) is dissolved in 0.6 mL of ethanol and diluted to 3 mL with sterile water. This solution is used to overlay a three day old

culture of *Saccharopolyspora erythraea* WHM34 (an *eryA* mutant) grown on a 100 mm R2YE agar plate at 30°C. After drying, the plate is incubated at 30°C for four days. The agar is chopped and then extracted three times with 100 mL portions of 1% triethylamine in ethyl acetate. The extracts are combined and evaporated. The crude product is purified by preparative HPLC (C-18 reversed phase, water-acetonitrile gradient containing 1% acetic acid). Fractions are analyzed by mass spectrometry, and those containing pure compound are pooled, neutralized with triethylamine, and evaporated to a syrup. The syrup is dissolved in water and extracted three times with equal volumes of ethyl acetate. The organic extracts are combined, washed once with saturated aqueous $NaHCO_3$, dried over Na_2SO_4, filtered, and evaporated to yield ~0.15 mg of product. The product is a glycosylated and hydroxylated oleandolide corresponding to erythromycin A, B, C, and D but differing therefrom as the oleandolide provided differed from 6-dEB.

Example 7

Measurement of Antibacterial Activity

Antibacterial activity is determined using either disk diffusion assays with *Bacillus cereus* as the test organism or by measurement of minimum inhibitory concentrations (MIC) in liquid culture against sensitive and resistant strains of *Staphylococcus pneumoniae*.

The invention having now been described by way of written description and example, those of skill in the art will recognize that the invention can be practiced in a variety of embodiments and that the foregoing description and examples are for purposes of illustration and not limitation of the following claims.

Claims

1. An isolated recombinant DNA compound that comprises a coding sequence for a domain of a loading module or any one of extender modules one through four, inclusive of an oleandolide polyketide synthase (PKS).

2. The isolated recombinant DNA compound of Claim 1, wherein said domain is selected from the group consisting of a thioesterase domain, a KS^Q domain, an AT domain, a KS domain, an ACP domain, a KR domain, a DH domain, and an ER domain.

3. The isolated recombinant DNA compound of Claim 2 that comprises the coding sequence for a loading module and extender modules one and two of the oleandolide PKS.

4. The isolated recombinant DNA compound of Claim 2 that comprises the coding sequence for the loading module and all six extender modules.

5. An isolated recombinant DNA compound that comprises a coding sequence for a domain of a loading module or any one of extender modules one through six, inclusive of an oleandolide polyketide synthase (PKS) operably linked to a promoter.

6. The isolated recombinant DNA compound of Claim 5, wherein said coding sequence encodes a loading module or any one of extender modules one through four, inclusive, of oleandolide PKS.

7. The isolated recombinant DNA compound of Claim 5 that is a recombinant DNA expression vector that further comprises an origin of replication or a segment of DNA that enables chromosomal integration.

8. The recombinant DNA expression vector of Claim 7 that codes for expression of a PKS in *Streptomyces* host cells.

9. A recombinant host cell selected from the group consisting of *Streptomyces* host cells and *Saccharopolyspora* host cells that comprises a recombinant DNA expression vector of Claim 7.

10. The recombinant DNA expression vector of Claim 7 that encodes a hybrid PKS comprising at least a portion of an oleandolide PKS gene and at least a portion of a second PKS gene for a macrolide aglycone other than oleandolide.

11. The recombinant DNA compound of Claim 10, wherein said second PKS gene is a DEBS gene.

12. The recombinant DNA compound of Claim 11, wherein said hybrid PKS comprises a loading module and any one of extender modules one through four, inclusive, of oleandolide PKS and an extender module of DEBS.

13. The recombinant DNA compound of Claim 10, wherein said hybrid PKS comprises a loading module and any one of extender modules one through four, inclusive, of oleandolide PKS and an extender module of narbonolide PKS.

14. A recombinant host cell, which in its untransformed state does not produce oleandolide, that comprises a recombinant DNA expression vector of Claim 11 and said cell produces a macrolide aglycone synthesized by said hybrid PKS.

15. The recombinant host cell of Claim 14 that is *Streptomyces lividans*.

16. The recombinant host cell of Claim 14 that is *Saccharopolyspora erythraea*.

17. The recombinant host cell of Claim 13, wherein said oleandolide PKS has a non-functional KS domain in extender module one.

18. The recombinant host cell of Claim 17 that is *Streptomyces coelicolor* or *Streptomyces lividans*.

19. The recombinant host cell of Claim 17 that is *Saccharopolyspora erythraea*.

20. A method for producing a polyketide in a cell, which method comprises transforming the cell with a recombinant expression vector that encodes at least a portion of an *oleAI*, *oleAII*, or *oleAIII* gene.

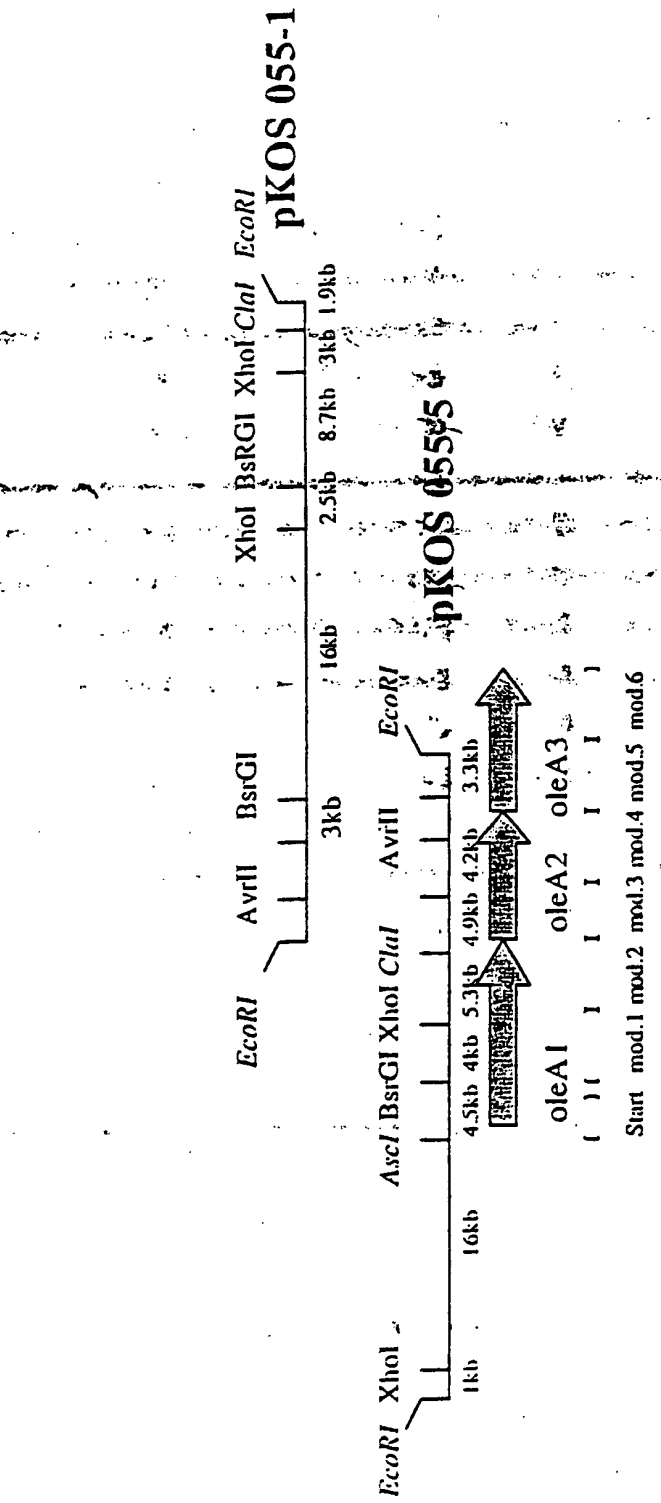

pKOS 055-1
EcoRI
AvrII
BsrGI
XhoI
BsrGI
XhoI
ClaI
EcoRI
3kb
16kb
2.5kb
8.7kb
3kb
1.9kb
pKOS 055-5
EcoRI
XhoI
AscI
BsrGI
XhoI
ClaI
AvrII
EcoRI
1kb
16kb
4.5kb
4kb
5.3kb
4.9kb
4.2kb
3.3kb
oleA1
oleA2
oleA3
Start mod.1 mod.2 mod.3 mod.4 mod.5 mod.6

Figure 1

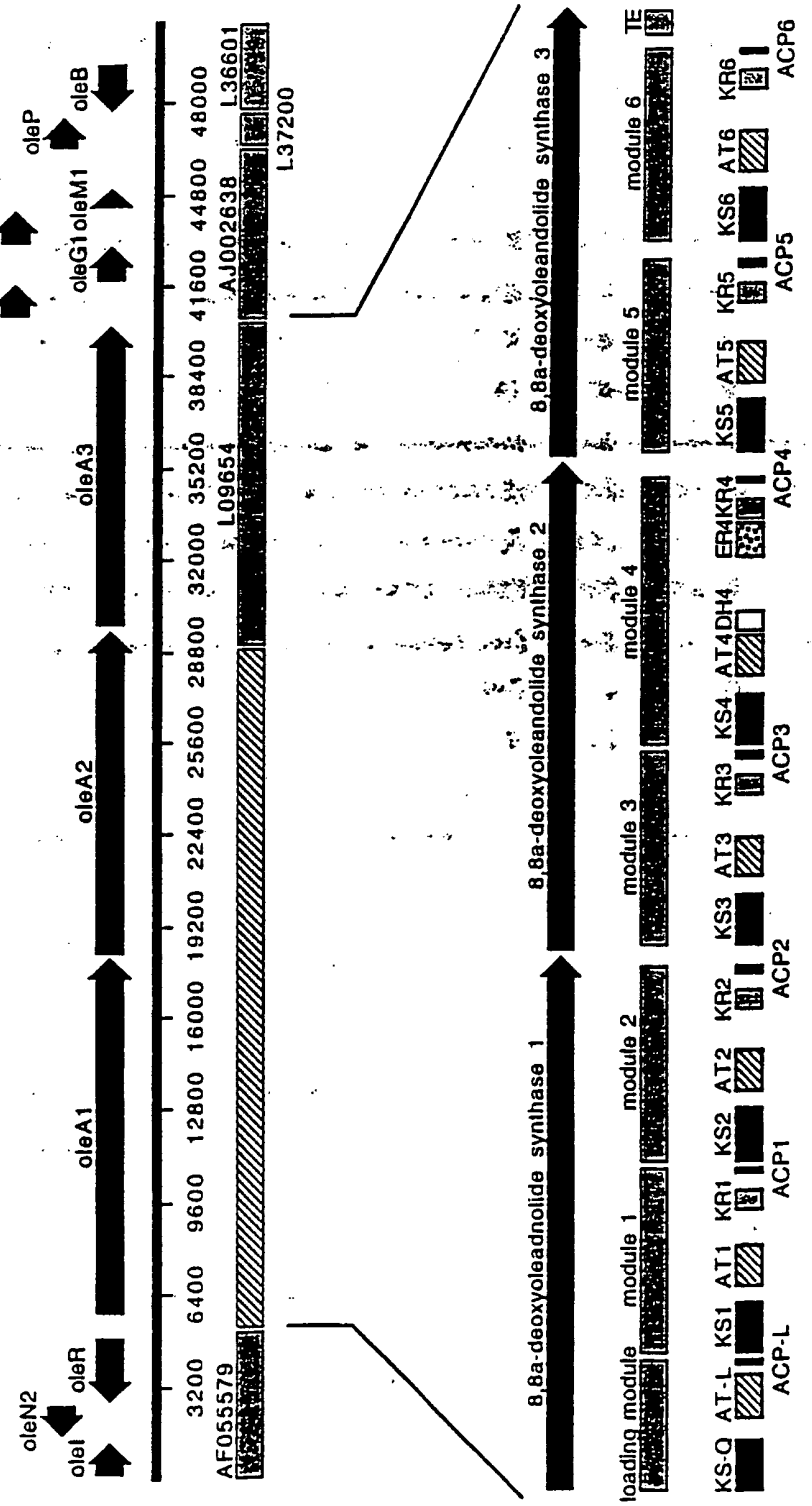
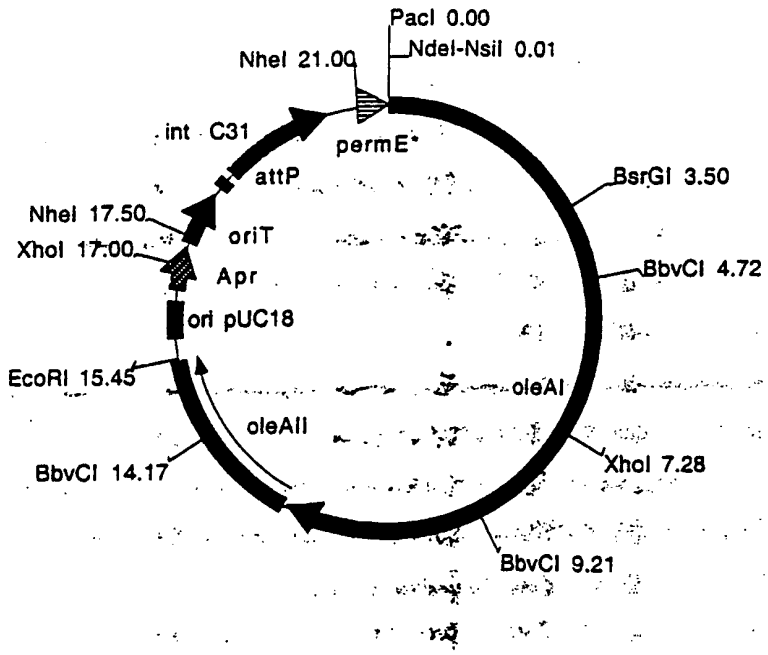

oleN2
oleI
oleR
oleA1
oleA2
oleA3
oleP1
oleG2
oleY
oleG1
oleM1
oleP
oleB
3200
6400
9600
12800
16000
19200
22400
25600
28800
32000
35200
38400
41600
44800
48000
AF055579
L09654
AJ002638
L37200
L36601
8,8a-deoxyoleadnolide synthase 1
8,8a-deoxyoleandolide synthase 2
8,8a-deoxyoleandolide synthase 3
loading module
module 1
module 2
module 3
module 4
module 5
module 6
TE
KS-Q
AT-L
ACP-L
KS1
AT1
KR1
ACP1
KS2
AT2
KR2
ACP2
KS3
AT3
KR3
ACP3
KS4
AT4
DH4
ER4
KR4
ACP4
KS5
AT5
KR5
ACP5
KS6
AT6
KR6
ACP6

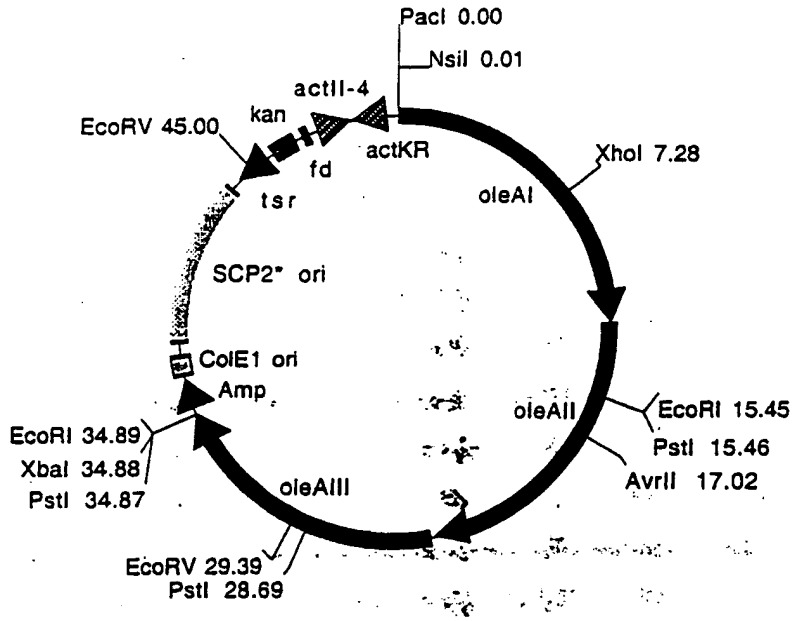
Figure 2



pKOS039-110

21.5 kb

Figure 3



pKOS039-130

51 kb

Figure 4

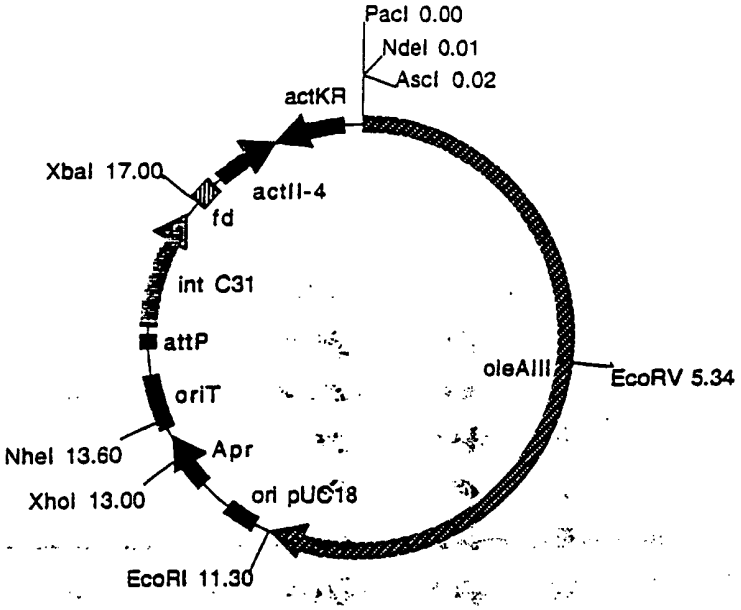

PacI 0.00
NdeI 0.01
AscI 0.02
actKR
XbaI 17.00
actII-4
fd
int C31
attP
oriT
oleAIII
EcoRV 5.34
NheI 13.60
Apr
XhoI 13.00
ori pUC18
EcoRI 11.30
pKOS039-133
19.8 kb

Figure 5

INTERNATIONAL SEARCH REPORT

Inter onal Application No
PCT/US 99/24478

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/54 C12P19/62 C12N15/63 C12N15/74 C12N15/62
C12N1/21 //(C12N1/21,C12R1:465),(C12N1/21,C12R1:01)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C12P

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 98 27203 A (KOSAN BIOSCIENCES) 25 June 1998 (1998-06-25) claim 1	20
Y	page 8, line 21,22 examples 5,6	1-16

[X] Further documents are listed in the continuation of box C. [X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
19 May 2000	20 JUNE 2000 (20.06.00)

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Herrmann, K

2

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Inte. onal Application No
PCT/US 99/24478

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
Y	SWAN DAVID G ET AL: "Characterisation of a Streptomyces antibioticus gene encoding a type I polyketide synthase which has an unusual coding sequence." MOLECULAR & GENERAL GENETICS 1994, vol. 242, no. 3, 1994, pages 358-362, XP002087278 ISSN: 0026-8925 abstract page 361, right-hand column, last paragraph & DATABASE EMBL 'Online! Accession No. L09654, 14 July 1994 (1994-07-14) "Streptomyces antibioticus polyketide synthase gene, complete cds of ORF3 subunit including module 5 and module 6" the whole document ---	1-16,20
Y	XUE YONGQUAN ET AL: "A gene cluster for macrolide antibiotic biosynthesis in Streptomyces venezuelae: Architecture of metabolic diversity." PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA OCT. 13; 1998, vol. 95, no. 21, 13 October 1998 (1998-10-13), pages 12111-12116, XP002138158 ISSN: 0027-8424 abstract & DATABASE EMBL 'Online! Accession No. AF079138, 22 October 1998 (1998-10-22) "Streptomyces venezuelae methymycin/pikromycin polykedide synthase gene cluster, complete sequence" the whole document --- -/--	1-16,20

2

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Inte onal Application No
PCT/US 99/24478

C.(Continuation) DOCUMENTS C NSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	OLANO C ET AL: "Analysis of a Streptomyces antibioticus chromosomal region involved in oleandomycin biosynthesis, which encodes two glycosyltransferases responsible for glycosylation of the macrolactone ring." MOLECULAR & GENERAL GENETICS AUG., 1998, vol. 259, no. 3, August 1998 (1998-08), pages 299-308, XP002096258 ISSN: 0026-8925 see page 14, line 4 of present description the whole document & DATABASE EMBL 'Online! Accession No. AJ002638, 1 October 1998 (1998-10-01) "Streptomyces antibioticus olePI, oleGI, oleMI and oleY genes" the whole document	1-20
T	TANG LI ET AL: "Formation of functional heterologous complexes using subunits from the picromycin, erythromycin and oleandomycin polyketide synthases." CHEMISTRY & BIOLOGY (LONDON) FEB., 2000, vol. 7, no. 2, February 2000 (2000-02), pages 77-84, XP000909347 ISSN: 1074-5521 the whole document	1-20

2

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 99/24478

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

see FURTHER INFORMATION sheet PCT/ISA/210

2. [ ] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

Intl. ional Application No
PCT/US 99/24478

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9827203 A	25-06-1998	AU 5701098 A	15-07-1998
		EP 0948613 A	13-10-1999